

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 30, 2001, 21:49:06 ; Search time 1167.42 Seconds
(without alignments)
7011.240 Million cell updates/sec

Title: US-09-382-242-23

Perfect score: 555
Sequence: 1 ATGCTTTTAAACAAGCAGCTC.....AATCATACACTTAACGATAG 555

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1283235 seqs, 7373929652 residues

Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*
1: gb_da1:*
2: gb_da2:*
3: gb_da3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_da1:*
17: em_da2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rnd:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rnd:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_v1:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_v11:*
59: gb_v12:*
60: gb_vtg1:*
61: gb_vtg2:*
62: gb_vtg3:*
63: gb_vtg4:*
64: gb_vtg5:*
65: gb_vtg6:*
66: gb_vtg7:*
67: gb_vtg8:*
68: gb_vtg9:*
69: gb_vtg10:*
70: gb_vtg11:*
71: gb_vtg12:*
72: gb_vtg13:*
73: gb_vtg14:*
74: gb_vtg15:*
75: gb_vtg16:*
76: gb_vtg17:*
77: gb_vtg18:*
78: gb_vtg19:*
79: gb_vtg20:*
80: gb_vtg21:*
81: gb_vtg22:*
82: gb_vtg23:*
83: gb_vtg24:*
84: gb_vtg25:*
85: gb_vtg26:*
86: gb_vtg27:*
87: gb_vtg28:*
88: gb_vtg29:*
89: gb_vtg30:*
90: gb_vtg31:*
91: gb_vtg32:*
92: gb_vtg33:*
93: gb_vtg34:*
94: gb_vtg35:*
95: gb_vtg36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	48.2	8.7	217412	69	AC0244400	AC0244400 Homo sapi
C 2	47.8	8.6	14001	6	PF00001R8	X95276 P. falciparum
C 3	47.6	8.6	192929	60	AC005505	AC005505 Plasmodium
C 4	47.6	8.6	256172	60	AC005139	AC005139 Plasmodium
C 5	47.6	8.6	310779	60	AC005140	AC005140 Plasmodium
C 6	47.2	8.5	108908	6	PF00001R8	AL034560 Plasmodium
C 7	47.2	8.5	298469	4	AE003846	AE003846 Drosophila
C 8	45.6	8.2	21595	64	AC014617	AC014617 Drosophila
C 9	45.6	8.2	193742	82	AL513530	AL513530 Homo sapi
C 10	45.2	8.1	3171	6	DD068754	U68754 Dictyostell
C 11	45	8.1	179401	62	AC010903	AC010903 Homo sapi

C 12	44.8	8.1	84346	67	AC021920	AC021920 Homo sapi
C 13	44.4	8.0	27694	71	HASMT	D31785 Pichia cana
C 14	44.4	8.0	185926	14	AC027632	AC027632 Homo sapi
C 15	44.2	8.0	118313	85	AC003991	AC003991 Human BAC
C 16	44.2	8.0	204652	84	PFM133P6	AL049183 Human BAC
C 17	44.4	7.9	1084	53	CNS071NH	AL425139 clone BAO
C 18	44.4	7.9	45565	8	FR15J11	AL021531 Fugu rubr
C 19	44.4	7.9	138094	70	AC026839	AC026839 Homo sapi
C 20	43.8	7.9	13859	4	AE001366	AC001366 Plasmodu
C 21	43.8	7.9	84346	67	PECOMPIRB	AC021920 Homo sapi
C 22	43.6	7.9	14001	6	PECOMPIRB	X52276 P. falciparu
C 23	43.6	7.9	15421	6	PECOMPIRB	X52275 P. falciparu
C 24	43.4	7.8	1847	5	AF209925	AF209925 Plasmodu
C 25	43.4	7.8	2787	6	PFSTARP	Z26314 P. falciparu
C 26	43.2	7.8	5708	13	AF271619	AF271619 Aspergill
C 27	43.2	7.8	16932	12	AF083031	AF083031 Guillard
C 28	43.2	7.8	169546	60	AC004157	AC004157 Plasmodu
C 29	43	7.7	34996	6	U87145	U87145 Toxoplasma
C 30	43	7.7	171798	4	AC084397	AC084397 Trypanoso
C 31	42.8	7.7	840	14	CNS0180K	AL110675 Bolyalis
C 32	42.8	7.7	149789	60	AC007926	AC007926 Trypanoso
C 33	42.8	7.7	149789	60	AC007926	AC007926 Trypanoso
C 34	42.8	7.7	180290	85	AC005081	AC005081 Homo sapi
C 35	42.6	7.7	1074	5	AF190049	AF190049 Pterostic
C 36	42.6	7.7	94384	87	AC011718	AC011718 Homo sapi
C 37	42.6	7.7	104992	60	AC005504	AC005504 Plasmodu
C 38	42.6	7.7	177951	76	AC074383	AC074383 Homo sapi
C 39	42.6	7.7	181980	63	AC013394	AC013394 Homo sapi
C 40	42.4	7.6	270	13	AY021751	AY021751 Oryza sat
C 41	42.2	7.6	801	53	CNS06EMJ	AL395297 T7 end of
C 42	42.2	7.6	1051	53	CNS06EMJ	AL392832 T3 end of
C 43	42.2	7.6	5769	5	AF104350	AF104350 Dictyoste
C 44	42.2	7.6	8212	43	ALR1BPTR	X75653 Astasia lon
C 45	42.2	7.6	8307	4	AC024860	AC024860 Caenorhab

ALIGNMENTS

RESULT 1	AC024400/c	217412 bp	DNA	HTG	13-JUL-2000
LOCUS	Homo sapiens clone Rp11-439d4				
DEFINITION	AC024400.2	GI:7767819			
ACCESSION	AC024400.2	GI:7767819			
VERSION	AC024400.2	GI:7767819			
KEYWORDS	human.				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1 (bases 1 to 217412)				
AUTHORS	Bliren, B., Linton, L., Nusbaum, C., Lander, E., Abrahams, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bede, F., Boguski, L., Bouck, G., Brown, A., Burkett, G., Campilano, A., Castle, A., Chong, Y., Collinge, M., Collins, S., Collins, A., Cooke, P., Dear, R., Dewar, K., Diaz, J., Dodge, S., Domingo, M., Doyle, M., Fagan, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J., Gardy, S., Ginde, S., Goyette, M., Graham, L., Grand, P., Grant, G., Hago, B., Heath, A., Horton, L., Howland, J., Iliev, I., Johnson, R., Jones, C., Kann, L., Karitas, A., Klein, J., Landers, T., Laroque, K., Lehoucq, J., Levine, R., Meneus, L., Mihov, T., Miranda, C., Mlepe, V., Morrow, J., Naylor, J., Norman, C., O'Connor, T., O'Donnell, P., O'Neill, D., Olyar, T.M., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testa, S., Theodore, J., Tittell, A.,				

JOURNAL

Travers, M., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On May 10, 2000 this sequence version replaced gi:7108401.

All repeats were identified using RepeatMasker:

http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

Project information

Center project name: l6776

Center clone name: 439_D_4

NOTE: This record contains 242 individual

sequencing reads that have not been assembled into

contigs. Runs of N are used to separate the reads

and the order in which they appear is completely

arbitrary. Low-pass sequence sampling is useful for

identifying clones that may be gene-rich and allows

overlap relationships among clones to be deduced.

However, it should not be assumed that this clone

will be sequenced to completion. In the event that

the record is updated, the accession number will

be preserved.

1 716 815: contig of 715 bp in length

816 1554: contig of 739 bp in length

1555 1654: gap of 100 bp

1655 2421: contig of 767 bp in length

2422 2521: gap of 100 bp

2522 3271: contig of 750 bp in length

3272 3371: gap of 100 bp

3372 4116: contig of 745 bp in length

4117 4216: gap of 100 bp

4217 4979: contig of 763 bp in length

4980 5079: gap of 100 bp

5080 5835: contig of 760 bp in length

5840 5939: gap of 100 bp

5940 6709: contig of 770 bp in length

6710 6809: gap of 100 bp

6810 7579: contig of 770 bp in length

7580 7679: gap of 100 bp

7680 8381: contig of 702 bp in length

8382 8481: gap of 100 bp

8482 9219: contig of 738 bp in length

9220 9319: gap of 100 bp

9320 10072: contig of 753 bp in length

10073 10172: gap of 100 bp

10173 10927: contig of 755 bp in length

10928 11027: gap of 100 bp

11028 11797: contig of 770 bp in length

11798 11897: gap of 100 bp

11898 12659: contig of 762 bp in length

12660 12759: gap of 100 bp

12760 13535: contig of 776 bp in length

13536 13635: gap of 100 bp

13636 14355: contig of 720 bp in length

14356 14455: gap of 100 bp

14456 15218: contig of 763 bp in length

15219 15319: gap of 100 bp

15319 16081: contig of 763 bp in length

16082 16181: gap of 100 bp

16182 16943: contig of 762 bp in length

16944 17043: gap of 100 bp

17044 17168: contig of 725 bp in length

17169 17669: gap of 100 bp

17669 18613: contig of 745 bp in length

*	18714	18713:	gap of	100	bp	in	length
*	18714	19460:	contig of	747	bp	in	length
*	19461	19560:	gap of	100	bp	in	length
*	19561	20328:	contig of	768	bp	in	length
*	20329	20428:	gap of	100	bp	in	length
*	20429	21151:	contig of	723	bp	in	length
*	21152	21251:	gap of	100	bp	in	length
*	21252	22017:	contig of	766	bp	in	length
*	22018	22117:	gap of	100	bp	in	length
*	22118	22872:	contig of	755	bp	in	length
*	22873	22972:	gap of	100	bp	in	length
*	22973	23743:	contig of	771	bp	in	length
*	23744	23843:	gap of	100	bp	in	length
*	23844	24584:	contig of	741	bp	in	length
*	24585	24684:	gap of	100	bp	in	length
*	24685	25443:	contig of	759	bp	in	length
*	25444	25543:	gap of	100	bp	in	length
*	25544	26270:	contig of	727	bp	in	length
*	26271	26370:	gap of	100	bp	in	length
*	26371	27124:	contig of	754	bp	in	length
*	27125	27224:	gap of	100	bp	in	length
*	27125	27980:	contig of	756	bp	in	length
*	27981	28080:	gap of	100	bp	in	length
*	28081	28811:	contig of	731	bp	in	length
*	28812	28911:	gap of	100	bp	in	length
*	28912	29655:	contig of	784	bp	in	length
*	29656	29795:	gap of	100	bp	in	length
*	29796	30568:	contig of	773	bp	in	length
*	30569	30668:	gap of	100	bp	in	length
*	30669	31438:	contig of	770	bp	in	length
*	31439	31538:	gap of	100	bp	in	length
*	31539	32218:	contig of	740	bp	in	length
*	32279	32378:	gap of	100	bp	in	length
*	33279	33108:	contig of	730	bp	in	length
*	33209	33208:	gap of	100	bp	in	length
*	33209	33965:	contig of	748	bp	in	length
*	33957	34056:	gap of	100	bp	in	length
*	34057	34680:	contig of	784	bp	in	length
*	34681	34940:	gap of	100	bp	in	length
*	34941	35674:	contig of	734	bp	in	length
*	35675	35774:	gap of	100	bp	in	length
*	35775	36534:	contig of	760	bp	in	length
*	36535	36634:	gap of	100	bp	in	length
*	36635	37400:	contig of	766	bp	in	length
*	37401	37500:	gap of	100	bp	in	length
*	37501	38289:	contig of	789	bp	in	length
*	38290	38389:	gap of	100	bp	in	length
*	38390	39147:	contig of	758	bp	in	length
*	39148	39247:	gap of	100	bp	in	length
*	39248	40007:	contig of	760	bp	in	length
*	40008	40107:	gap of	100	bp	in	length
*	40108	40835:	contig of	728	bp	in	length
*	40836	40935:	gap of	100	bp	in	length
*	40936	41677:	contig of	742	bp	in	length
*	41678	41777:	gap of	100	bp	in	length
*	41778	42547:	contig of	770	bp	in	length
*	42548	42647:	gap of	100	bp	in	length
*	42648	43385:	contig of	738	bp	in	length
*	43386	43485:	gap of	100	bp	in	length
*	43486	44251:	contig of	766	bp	in	length
*	44252	44351:	gap of	100	bp	in	length
*	44352	45108:	contig of	757	bp	in	length
*	45109	45208:	gap of	100	bp	in	length
*	45209	45973:	contig of	765	bp	in	length
*	45974	46073:	gap of	100	bp	in	length
*	46074	46842:	contig of	769	bp	in	length
*	46843	46942:	gap of				

*	496.98	50197:	contlg of 730 bp	in length
*	501.98	50297:	gap of 100 bp	
*	502.98	51049:	contlg of 752 bp	in length
*	510.50	51149:	gap of 100 bp	
*	511.50	51907:	contlg of 758 bp	in length
*	513.08	52007:	gap of 100 bp	
*	520.08	52728:	contlg of 721 bp	in length
*	527.22	52898:	gap of 100 bp	
*	528.22	53594:	contlg of 766 bp	in length
*	530.95	53694:	gap of 100 bp	
*	533.95	54432:	contlg of 738 bp	in length
*	536.95	54532:	gap of 100 bp	
*	544.33	54532:	gap of 100 bp	
*	545.33	55255:	contlg of 763 bp	in length
*	552.55	55395:	gap of 100 bp	
*	553.96	56159:	contlg of 764 bp	in length
*	561.60	56299:	gap of 100 bp	
*	562.60	57028:	contlg of 769 bp	in length
*	570.29	57128:	gap of 100 bp	
*	571.29	57851:	contlg of 723 bp	in length
*	578.52	57951:	gap of 100 bp	
*	579.52	58690:	contlg of 739 bp	in length
*	586.91	58790:	gap of 100 bp	

Query Match	8.7%;	Score 48.2;	DB 69;	Length 217412;
Best Local Similarity	48.6%;	Pred. No. 1.5;		
Matches 160;	Conservative 0;	Mismatches 168;	Indels 1;	Gaps 1

122 CATTAGCGATCTAGGACATGCTGTTAAAGCAGTGTGCTCCATATTCATCTAGGC 181

Db 62840 CATGTATGATTCAGGATATCTTCATAATATATGTTCCGCATATATTCATGATATATA 62781

182 TGGCAATGGGTGGGATACCTAAATGGTTATAGTGGGTAAAGAAATCTTATATTCGTGGAGTA 241

[illegible][illegible]

2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030 2031 2032 2033 2034 2035 2036 2037 2038 2039 2040 2041 2042 2043 2044 2045 2046 2047 2048 2049 2050 2051 2052 2053 2054 2055 2056 2057 2058 2059 2060 2061 2062 2063 2064 2065 2066 2067 2068 2069 2070 2071 2072 2073 2074 2075 2076 2077 2078 2079 2080 2081 2082 2083 2084 2085 2086 2087 2088 2089 2090 2091 2092 2093 2094 2095 2096 2097 2098 2099 2100 2101 2102 2103 2104 2105 2106 2107 2108 2109 2110 2111 2112 2113 2114 2115 2116 2117 2118 2119 2120 2121 2122 2123 2124 2125 2126 2127 2128 2129 2130 2131 2132 2133 2134 2135 2136 2137 2138 2139 2140 2141 2142 2143 2144 2145 2146 2147 2148 2149 2150 2151 2152 2153 2154 2155 2156 2157 2158 2159 2160 2161 2162 2163 2164 2165 2166 2167 2168 2169 2170 2171 2172 2173 2174 2175 2176 2177 2178 2179 2180 2181 2182 2183 2184 2185 2186 2187 2188 2189 2190 2191 2192 2193 2194 2195 2196 2197 2198 2199 2200 2201 2202 2203 2204 2205 2206 2207 2208 2209 2210 2211 2212 2213 2214 2215 2216 2217 2218 2219 2220 2221 2222 2223 2224 2225 2226 2227 2228 2229 2230 2231 2232 2233 2234 2235 2236 2237 2238 2239 2240 2241 2242 2243 2244 2245 2246 2247 2248 2249 2250 2251 2252 2253 2254 2255 2256 2257 2258 2259 2260 2261 2262 2263 2264 2265 2266 2267 2268 2269 2270 2271 2272 2273 2274 2275 2276 2277 2278 2279 2280 2281 2282 2283 2284 2285 2286 2287 2288 2289 2290 2291 2292 2293 2294 2295 2296 2297 2298 2299 2300 2301 2302 2303 2304 2305 2306 2307 2308 2309 2310 2311 2312 2313 2314 2315 2316 2317 2318 2319 2320 2321 2322 2323 2324 2325 2326 2327 2328 2329 2330 2331 2332 2333 2334 2335 2336 2337 2338 2339 2340 2341 2342 2343 2344 2345 2346 2347 2348 2349 2350 2351 2352 2353 2354 2355 2356 2357 2358 2359 2360 2361 2362 2363 2364 2365 2366 2367 2368 2369 2370 2371 2372 2373 2374 2375 2376 2377 2378 2379 2380 2381 2382 2383 2384 2385 2386 2387 2388 2389 2390 2391 2392 2393 2394 2395 2396 2397 2398 2399 2400 2401 2402 2403 2404 2405 2406 2407 2408 2409 2410 2411 2412 2413 2414 2415 2416 2417 2418 2419 2420 2421 2422 2423 2424 2425 2426 2427 2428 2429 2430 2431 2432 2433 2434 2435 2436 2437 2438 2439 2440 2441 2442 2443 2444 2445 2446 2447 2448 2449 2450 2451 2452 2453 2454 2455 2456 2457 2458 2459 2460 2461 2462 2463 2464 2465 2466 2467 2468 2469 2470 2471 2472 2473 2474 2475 2476 2477 2478 2479 2480 2481 2482 2483 2484 2485 2486 2487 2488 2489 2490 2491 2492 2493 2494 2495 2496 2497 2498 2499 2500 2501 2502 2503 2504 2505 2506 2507 2508 2509 2510 2511 2512 2513 2514 2515 2516 2517 2518 2519 2520 2521 2522 2523 2524 2525 2526 2527 2528 2529 2530 2531 2532 2533 2534 2535 2536 2537 2538 2539 2540 2541 2542 2543 2544 2545 2546 2547 2548 2549 2550 2551 2552 2553 2554 2555 2556 2557 2558 2559 2560 2561 2562 2563 2564 2565 2566 2567 2568 2569 2570 2571 2572 2573 2574 2575 2576 2577 2578 2579 2580 2581 2582 2583 2584 2585 2586 2587 2588 2589 2590 2591 2592 2593 2594 2595 2596 2597 2598 2599 2600 2601 2602 2603 2604 2605 2606 2607 2608 2609 2610 2611 2612 2613 2614 2615 2616 2617 2618 2619 2620 2621 2622 2623 2624 2625 2626 2627 2628 2629 2630 2631 2632 2633 2634 2635 2636 2637 2638 2639 2640 2641 2642 2643 2644 2645 2646 2647 2648 2649 2650 2651 2652 2653 2654 2655 2656 2657 2658 2659 2660 2661 2662 2663 2664 2665 2666 2667 2668 2669 2670 2671 2672 2673 2674 2675 2676 2677 2678 2679 2680 2681 2682 2683 2684 2685 2686 2687 2688 2689 2690 2691 2692 2693 2694 2695 2696 2697 2698 2699 2700 2701 2702 2703 2704 2705 2706 2707 2708 2709 2710 2711 2712 2713 2714 2715 2716 2717 2718 2719 2720 2721 2722 2723 2724 2725 2726 2727 2728 2729 2730 2731 2732 2733 2734 2735 2736 2737 2738 2739 2740 2741 2742 2743 2744 2745 2746 2747 2748 2749 2750 2751 2752 2753 2754 2755 2756 2757 2758 2759 2760 2761 2762 2763 2764 2765 2766 2767 2768 2769 2770 2771 2772 2773 2774 2775 2776 2777 2778 2779 2780 2781 2782 2783 2784 2785 2786 2787 2788 2789 2790 2791 2792 2793 2794 2795 2796 2797 2798 2799 2800 2801 2802 2803 2804 2805 2806 2807 2808 2809 2810 2811 2812 2813 2814 2815 2816 2817 2818

[illegible]

361 GCATATTCATAGCTATATCAATACCTTGACCAATCATGGATAGCTGTTCTACTAATAATA 420

Db 62600 TATATATCATATATATTCATGATAGTATATTCATATATATATATTCATAATATATAT 62541

0Y 421 GGTCAATTTGCAATGTGGTATCTACACTT 449

Db 62540 TCATATATATTCATAATATATATTCAT 62512

РЕСПУБЛИКА

PECOMPTRB/c		
PECOMPTRB	14001	bn
PECOMPTRB		DNA
PECOMPTRB		INV
PECOMPTRB		14-FEB-1997

DEFINITION	P. taicaiiparum complete gene map of plastid-like DNA (LR-B).
ACCESSION	X95276

VERSION	Accession
KEYWORDS	Clp gene; LSU rRNA gene; ORF105; ORF129; ORF78; ORF79; ORF91; rRNA

rpl6 gene; rps11 gene; rps12 gene; rps17 gene; rps19 gene; rps3

tRNA-Cys; tRNA-Gln; tRNA-Glu; tRNA-Gly; tRNA-His; tRNA-Leu;

SOURCE
malaria parasite P. falciparum
tRNA-Trp; tRNA-Tyr; tufa gene.

ORGANISM	
Plasmodium talcifarum	
Eukaryota: Alveolata: Apicomplexa: Plasmodium	

REFERENCE	1 - (bases 1 to 14001)
AUTHORS	Wilson, R.J.M., Denny, P.W., Preiser, P.R., Rangachari, K., Roberts,

Williamson, D.H.

TITLE	WATKINSON/D.M.
Complete gene map of the plastid-like DNA of the malaria parasite	

Plasmodium falciparum
J. Mol. Biol. 261 (2), 155-172 (1996)
65346169
REFERENCE 2 (bases 1 to 14001)
AUTHORS Wilson, R.J.M.
TITLE Direct Submission
JOURNAL Submitted (23-JAN-1996) R.J.M. Wilson, National Institute for
Medical Research, Mill Hill, London NW7 1AA, UK
COMMENT Related sequences X90351-X90354 (Incl.), and X87630-X87631 (Incl.).
FEATURES
source
1. 14001
/organism="Plasmodium falciparum"
/strain="C10"
/db_xref="taxon:5833"
/dev_stage="erythrocytic"
/note="IR-B half of 35kb circle, putative plastid DNA"
<1. 227
/gene="LSU rRNA"
/product="large subunit ribosomal RNA"
1. 227
/gene="LSU rRNA"
242. 314
/gene="tRNA-Thr"
/product="transfer RNA-Thr"
242. 314
/gene="tRNA-Thr"
340. 966
/gene="rps4"
340. 966
/gene="rps4"
/codon_start=1
/protein_id="CAA64576.1"
/db_xref="GI:1171592"
/db_xref="SPTREMBL:Q25804"
/translation="MIKELKPKIKILKLNIPFLYLSGKYNYKLNKIKSYKSEFDL
KLKPIRYGVNCTYKLYLKNKIDKNNINILYFKLKLLELDLDELINIGPEFT
ILDSRYIKYKNIYINININKYNNINLNKNDILFPNNKIKITILLKNIYKNIYIYS
NLRYNFKITSYKPFILCIYKFKILNINLNILYINDIYI"
981. 1053
/gene="tRNA-His"
/product="transfer RNA-His"
981. 1053
/gene="tRNA-His"
1072. 1142
/gene="tRNA-Cys"
/product="transfer RNA-Cys"
1072. 1142
/gene="tRNA-Cys"
1148. 1183
/gene="tRNA-Cys"
1148. 1183
/gene="tRNA-Leu"
/note="extein 1"
/product="transfer RNA-Leu"
1148. 1183
/gene="tRNA-Leu"
1184. 1317
/gene="tRNA-Leu"
1184. 1317
/gene="tRNA-Leu"
1318. 1366
/gene="tRNA-Leu"
1318. 1366
/gene="tRNA-Met (1)"
/product="transfer RNA-Met (1)"
1373. 1452
/gene="tRNA-Met (1)"
1457. 1541
/gene="tRNA-Tyr"
/product="transfer RNA-Tyr"
1457. 1541
gene

trna
1555. 1646
/gene="tRNA-Tyr"
1555. 1646
/gene="tRNA-Ser"
/product="transfer RNA-Ser"
1555. 1646
/gene="tRNA-Ser"
1653. 1728
/gene="tRNA-Asp"
/product="transfer RNA-Asp"
1653. 1728
/gene="tRNA-Asp"
1739. 1811
/gene="tRNA-Lys"
1739. 1811
/gene="tRNA-Lys"
1819. 1888
/gene="tRNA-Glu"
/product="transfer RNA-Glu"
1819. 1888
/gene="tRNA-Glu"
1901. 1972
/gene="tRNA-Pro"
/product="transfer RNA-Pro"
1901. 1972
/gene="tRNA-Pro"
1999. 2571
/gene="rpl4"
1999. 2571
/gene="rpl4"
/gene="rpl4"
/codon_start=1
/protein_id="CAA64577.1"
/db_xref="GI:1171593"
/db_xref="SPTREMBL:Q25805"
/translation="MNIIILNNNLNIIIEFKYKYNFPIKLYFNNTIKIKLIYIYK
LYINITYMKYTKRKSQVPSNKRIRQKIGKARLNKSPKQACACNPGPYKKN
KIISKINRLFLVLLNKRNSITITILENTINILNIPYKKNYCIFFKLGLGIIINL
KYILINLNKLFNNIFINIMYVLIPLI"
2575. 2802
/gene="rpl23"
2575. 2802
/gene="rpl23"
/codon_start=1
/protein_id="CAA64578.1"
/db_xref="GI:1171594"
/db_xref="SPTREMBL:Q25806"
/translation="MKVILNFEYLYNLFYKINVLNKFCLIIYSIKYFTKIDIKYIKN
IFKTLINNNIKINNINKKCKLKYITTR"
2799. 3536
/gene="rpl2"
2799. 3536
/gene="rpl2"
/codon_start=1
/protein_id="CAA64579.1"
/db_xref="GI:1171595"
/db_xref="SPTREMBL:Q25807"
/translation="MILKLYNTYKFKSGKNNKGYITLYNKGGNLKYNYKLIDS
WYDNYNTITINYKVLFFKIKIKYFNRYIGCIIYLSKLNLOKFTIILQHNIMNSIYVL
TNINIKTGSYIOLKCYKLGTYIYNISQYKKSIFARSAGTFQOLISFYKNLYIKL
PSKCKYININNFCYIGVNSINIEHNKIKINAGNINITYINIKPVRGAKAVCDHPHOG
GKCKTGIGRKYPCCKKGLHSKGYTIR"
3537. 3829
/gene="rps19"
3537. 3829
/gene="rps19"
/codon_start=1
/protein_id="CAA64580.1"
/db_xref="GI:1171596"
/db_xref="SPTREMBL:Q25808"
/translation="MIKLYMLKITINNKYIIFININKYKYNKLNILNINKNLYIYKTL
LNITYKYNCKKFPYIYINKTKLENKGNFIYTFVKNNIKELLIN"
3839. 4483
/gene="rps3"
gene

Db 192255 AACATTTTTCCTTTCTTTTATTTATGATATATATATTTTAAACATTTT 192196
 Oy 454 TCAGAGATTCGAGAGGTCGTCATTCGCCAGATATAGGGATTCGCTTTTACCA 513
 Db 192195 TTTTATTTAAATAATTTGTTTATTTATTTCTAATAAATTTTATATATATAAT 192136
 Oy 514 TTCGAT 519
 Db 192135 TTTT 192130

RESULT 4
 AC005139/c DNA HTG 01-APR-1999
 LOCUS Plasmidium falciparum chromosome 12, *** SEQUENCING IN PROGRESS
 DEFINITION *** 5 unordered pieces.
 ACCESSION AC005139 GI:4558581
 VERSION AC005139.3
 KEYWORDS HTG; HTGS_PHASE1.
 SOURCE malaria parasite P. falciparum.
 ORGANISM Plasmidium falciparum
 Eukaryote; Alveolata; Apicomplexa; Haemosporidia; Plasmidium.
 REFERENCE 1 (bases 1 to 256172)
 AUTHORS Hyman,R.W., Fung,E.L., Qin,F., Tamaki,T., Kurdi,O.B. and Davis,R.W.
 TITLE Plasmidium falciparum 3D7 chromosome 12
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 256172)
 AUTHORS Hyman,R.W., Fung,E.L., Qin,F., Tamaki,T., Kurdi,O.B. and Davis,R.W.
 TITLE Direct Submission
 JOURNAL Submitted (18-JUN-1998) Stanford DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA

COMMENT
 On Apr 2, 1999 this sequence version replaced gi:4331170.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 5 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 126535: contig of 126535 bp in length
 * 126536 126735: gap of unknown length
 * 126736 158845: contig of 32110 bp in length
 * 158846 159045: gap of unknown length
 * 159046 211538: contig of 52493 bp in length
 * 211539 211738: gap of unknown length
 * 211739 248106: contig of 36366 bp in length
 * 248107 248306: gap of unknown length
 * 248307 256172: contig of 7866 bp in length.
 Location/Qualifiers
 1. 256172
 /organism="Plasmidium falciparum"
 /db_xref="taxon:5833"
 /chromosome="12"

BASE COUNT 100700 a 26900 c 25826 g 101937 t 809 others
 ORIGIN

Query Match 8.68; Score 47.6; DB 60; Length 256172;
 Best Local Similarity 43.68; Pred. No. 2;
 Matches 212; Conservative 0; Mismatches 274; Indels 0; Gaps 0;

Oy 34 ATAAATATTTATTCAGCTTTCTTCCATTAACAATGATCGATATCTCATG 93
 Db 145961 ATAAATTTCCACAGATGATTTGTTTTTTCTTAAAAAAAACACATATATCTGATCTA 145902
 Oy 94 TCGTCATGTTTATATATGGAATATGATTAAGCATCTAGACATGCTGTTAAAGC 153
 Db 145901 TATATATGATATTTAGTATATATTTAGTATGATGATGATGATATATATA 145842
 Oy 154 AGTGTGCTCAATATTCATCTAGCTTTCATTTGCTGGGATACATATTTGTTATAGT 213

Db 145841 TATGATGCTAAATGCAATGATGATATATATATATATATATGTAATGTC 145782
 Oy 214 GGTTAACAATTCCTTATTCGTGAGTAGATTAAGATCTTTAATCATATCATGGGT 273
 Db 145781 AATGATTAATTTTATTTTATTTTATTTTATTTTAAATTTTATTTTATTTT 145722
 Oy 274 GTATTTCTTAACCTTAATAGGGGTTTTCGACGAGATATAGTGGATACATCTCATGTC 333
 Db 145721 ATTTAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 145662
 Oy 334 TCAGTATGTTTTCTTATTCATATATACCAATATTTATGCTATATCATCTTGACAAA 393
 Db 145661 ATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 145602
 Oy 394 TCATGATAGCTGTTCTCTAATAATAGTCAATATGCAATGCTATCTACATTTGCT 453
 Db 145601 AACATTTTTCCTTTCTTTTATTTATGATATATATATTTTATTTTATTTTATTTT 145542
 Oy 454 TCAGAGATTCGAGAGGTCGTCATTCGCCAGATTAATAGCGGATTCGTTTTTACCA 513
 Db 145541 TTTTATTTAATAATTTGTTTATTTATTTATTTCTAATAAATTTTATATATATAAT 145482
 Oy 514 TTCGAT 519
 Db 145481 TTTT 145476

RESULT 5
 AC005140/c DNA HTG 12-AUG-2000
 LOCUS Plasmidium falciparum chromosome 12 clone 3D7, *** SEQUENCING IN
 DEFINITION PROGRESS *** 4 unordered pieces.
 ACCESSION AC005140 GI:9797735
 VERSION AC005140.8
 KEYWORDS HTG; HTGS_PHASE1.
 SOURCE malaria parasite P. falciparum.
 ORGANISM Plasmidium falciparum
 Eukaryote; Alveolata; Apicomplexa; Haemosporidia; Plasmidium.
 REFERENCE 1 (bases 1 to 310779)
 AUTHORS Hyman,R.W., Fung,E.L., Qin,F., Rowley,D., Mao,J., Tamaki,T.,
 Kurdi,O.B., Conway,A.B. and Davis,R.W.
 TITLE Plasmidium falciparum 3D7 chromosome 12
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 310779)
 AUTHORS Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.
 TITLE Direct Submission
 JOURNAL Submitted (18-JUN-1998) Stanford DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA

COMMENT
 On Aug 12, 2000 this sequence version replaced gi:8810455.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 4 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 187308: contig of 187308 bp in length
 * 187309 187508: gap of unknown length
 * 187509 257820: contig of 70312 bp in length
 * 257821 258020: gap of unknown length
 * 258021 307255: contig of 49235 bp in length
 * 307256 307455: gap of unknown length
 * 307456 310779: contig of 3324 bp in length.
 Location/Qualifiers
 1. 310779
 /organism="Plasmidium falciparum"
 /db_xref="taxon:5833"
 /chromosome="12"
 /clone="PRYACB8-420"
 /clone="3D7"

FEATURES
 source

SEYENDINDIYVPGSKYKTLIEVLEPSCNNTPASGKNTPSDTQNDIONDGPSTP
PNTSPDIPPSDIPPTIDDEWNTLDEFIISOLSEOPNDPNDKSNSTNTNI
TTTSHNVEKPEPIMSJHBRDLYGEBYENNSNGIYSSNSRDLSCGTKYVSGI
DLINDLSNGNOPIDIDEILKREKENLFGTNNKOTSHSAKLNSDIDJONLEF
HKMLDRHRCCEKLNENRKLKEMEETHSGNTHSDSNKTLNTDSDIOMD
GKREKPTNMDTLEDKDKNEPRYDVODDIYDVNDHDVSTVDSNMDIPSKVQIEM
DVNTKLVKEKPEPIADVMT
38960..38965
/note="corrected splice donor for PFC0005c"
42416..43015
/note="R-PA3"
complement(join(43045..43971,44202..44255))
/gene="PFC0010c, MAL3P8.2"
complement(join(43045..43971,44202..44255))
/gene="PFC0010c, MAL3P8.2"
/note="PFC0010c (MAL3P8.2), r1fin (3D7-r1fT3-1), len: 326
aa: Similarity to other P.falciptarum r1fins. Member of a
multi-gene family; predicted using hexexon"
/codon_start=1
/protein_id="CAB39116.1"
/db_xref="GI:4494021"
/translation="MKIHYINILFLPLNLYNORNNHYITRPKATRTICECELY
APSNVSDPEMKRVMOQFVDRTORFHEVDERMKTRPKKCEQDCKEIOKILKDKLE
KMAOOLUTPDNITPTEPTICERSLAKTERPCINCRTMGVAPRGVLGCLCY
VGTNVTYQTAOKGIEAGYKAGIEBKGPGLKLIKVSQIOSFINPANAEBETTF
NFKOVENKCVKKEHFECSYTSIKENALSORASETAKNAADMAKLAEEVLDKAP
VTSGLNTAIIASIVAILVLMIIYFLRYRKNKKKKKLYIKLEE"
complement(46448..47828)
/gene="PFC0015c"
complement(46448..47828)
/gene="PFC0015c"
/note="PFC0015c varc pseudogene"
/codon_start=1
/pseudo
complement(join(48956..49777,49881..49949))
/gene="PFC0025c, MAL3P8.4"
complement(join(48956..49777,49881..49949))
/gene="PFC0025c, MAL3P8.4"
/note="PFC0025c (MAL3P8.4), stavor (3D7-stavorT3-1), len:
296 aa: Similarity to other P.falciptarum stavors. Member
of a multi-gene family; predicted using hexexon"
/codon_start=1
/protein_id="CAB39117.1"
/db_xref="GI:4494022"
/translation="MKMYLKMFLFTFLINTLIPHYENLYNDYNSFYONNTKRT
IKSRLLAQTOIHNPYHNDPELKEIIDKMEBAIKRYOCHDPEYQLEKVEYKNGTXY
SGNDAPETLLEKELLEYEIEFGNSDMLKSGMNVDERKSTFECQDINCYNLAK
TKGRDYIKHLKRGCTGYFCGSLGVSILTYICTTAAGKAISTAGAGSKNCYATSI
FHLNKKSSMSATFALCTGVCADYASGATSAATFPCIAALVLLITLVLLITLI
WLRKRKSKWKECKHCK"complement(join(52066..53076,53192..53260))
/gene="PFC0030c, MAL3P8.5"
complement(join(52066..53076,53192..53260))
/gene="PFC0030c, MAL3P8.5"
/note="PFC0030c (MAL3P8.5), r1fin (3D7-r1fT3-2), len: 359
aa: Similarity to other P.falciptarum r1fins. Member of a
multi-gene family; predicted using hexexon"
/codon_start=1
/protein_id="CAB39118.1"
/db_xref="GI:4494023"
/translation="MKLIYSNILLFALANLILTYVANKNKPSTPHHTRTTSL
LSCDTSSIYSDDEINYSKEIFEROTSRFEYERMTTORKEADNDKIOIKI
HKDMKNLAEKIEKGLRCGGLGASGIVGAVAIIMRPALDAIAKAEI
TANISAGVAGATGKVLIVISLKMGISLGLKNLESFATTSYKVKVIAVVE
OHEKCEPGYKGFPTVGPDSRDIHRCOSWMOGTSVSTGYISPKDVKIKTVE
VSAEBPANAFAEVKATERATIKAAEKEKITEASQLSAIGSILAILIIVLMIL
ITILYRKRKKKKKKAQYTKLINE"
complement(join(58103..58174,58338..59390))
/gene="PFC0035w, MAL3P8.6"
join(58103..58174,58338..59390)
/gene="PFC0035w, MAL3P8.6"

Query Match 8.58; Score 47.2; DB 6; Length 108908;
Best Local Similarity 47.98; Pred. No. 2.5;
Matches 136; Conservative 0; Mismatches 148; Indels 0; Gaps 0;
144 TGTAAAGCAGTGTGTCATATTCATCTAGCTTCATGTCGATGCGGATCATAT 203
Db 73231 TATTAATRAAAAATATATTTTATAGTTTTTAAAGATTTTCATTTGTGAAAGGACAT 73172
QY 204 TGTATAGTTGGTTTAAAGAAATCTTATTCGTCGATGAGTAAAGATCTTAAATCAT 263
Db 73171 TTTAATATTTTATTTAAGAACATATGTTTAAAGATATTTCTTGTATTTTATATAGT 73112
QY 264 ATCCATGGCTATTTCTTACTTATTAAGCGTTTTCGACGAAGTATATGTTGATACACA 323
Db 73111 AATATAGCATGTATATTAAGAATTAATGTTTCTACACAGTACATATGTTCTGTTT 73052
QY 324 TTTCTAGTCTCAGTATGTTTCTTATCATATATGACATATTCATAGCATATATCAT 383
Db 73051 TTGAATATTTACTATAGTTTATATATTTTCTTAAAGCATCTTTTGGATATAT 72992
QY 384 ACTTGACAATCATGATAGCTGTTCTACTAATATATAGTCATA 427
Db 72991 ACATACATACATATATATATATTTATTAAGAAGGTGATA 72948
RESULT 7
AE003846
LOCUS 298469 bp DNA INV 04-OCT-2000
DEFINITION Drosophila melanogaster genomic scaffold 14200001338051 section 1
ACCESSION AE003846 AE002804
VERSION AE003846.2 GI:10728138
KEYWORDS HMG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachyera;
Muscophora; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 298469)
Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.G., Mortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
Brannon,R.C., Rogers,J.H., Blazek,R.G., Chample,M., Pfeiffer,B.D.,
/note="PFC0035w (MAL3P8.6), r1fin (3D7-r1fT3-3), len: 375
aa: Similarity to other P.falciptarum r1fins. Member of a
multi-gene family; predicted using hexexon"
/codon_start=1
/protein_id="CAB39119.1"
/db_xref="GI:4494024"
/translation="MRITIKMKVYINILALSLNLTANTLANTHOKSPSPRHIOITRL
CECELYNHNNDPEMKRVMOQFVDRTORFHEVDERMKTRPKKCEQDCKEIOKILKDKLE
KMAOOLUTPDNITPTEPTICERSLAKTERPCINCRTMGVAPRGVLGCLCY
VGTNVTYQTAOKGIEAGYKAGIEBKGPGLKLIKVSQIOSFINPANAEBETTF
NFKOVENKCVKKEHFECSYTSIKENALSORASETAKNAADMAKLAEEVLDKAP
VTSGLNTAIIASIVAILVLMIIYFLRYRKNKKKKKLYIKLEE"
complement(61248..61301,61548..62459)
/gene="PFC0040w, MAL3P8.7"
join(61248..61301,61548..62459)
/gene="PFC0040w, MAL3P8.7"
/note="PFC0040w (MAL3P8.7), r1fin (3D7-r1fT3-4), len: 322
aa: Similarity to other P.falciptarum r1fins. Member of a
multi-gene family; predicted using hexexon"
/codon_start=1
/protein_id="CAB39120.1"
/db_xref="GI:4494025"
/translation="MKIHYINILFLPLNLYNORNNHYITRPKATRTICECELY
APSNVSDPEMKRVMOQFVDRTORFHEVDERMKTRPKKCEQDCKEIOKILKDKLE
KMAOOLUTPDNITPTEPTICERSLAKTERPCINCRTMGVAPRGVLGCLCY
VGTNVTYQTAOKGIEAGYKAGIEBKGPGLKLIKVSQIOSFINPANAEBETTF
NFKOVENKCVKKEHFECSYTSIKENALSORASETAKNAADMAKLAEEVLDKAP
VTSGLNTAIIASIVAILVLMIIYFLRYRKNKKKKKLYIKLEE"

Man, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor
Miklos, G.L., Abri, J.F., Agbayan, A., An, H.J.,
Andrews-Ffiankooch, C., Baldwin, D., Bailew, R.M., Basu, A.,
Bendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y.,
Benson, P.V., Berman, B.P., Bhandari, D., Bolshakov, S., Borkova, D.,
Botchan, M.R., Bock, J., Brokstein, P., Brotlier, P., Burtis, K.C.,
Busam, D.A., Butler, H., Cadieu, E., Cavenot, A., Chandra, I.,
Cherry, J.M., Cawley, S., Dahke, C., Davenport, L.B., Davies, P., de
Pablo, B., Delcher, A., Deng, Z., Mays, A.D., Dev, I., Dietz, S.M.,
Dodson, K., Dou, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C.,
Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferriera, S.,
Fleischmann, W., Foster, C., Gabrielian, A.E., Garg, N.S.,
Gelbart, M.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z.,
Guan, P., Harris, M., Harris, N.L., Harvey, D., Helman, T.J.,
Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J.,
Wei, M.H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z.,
Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C.,
Krautz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A.,
Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C.,
McLeod, M.P., McPherson, D., Merkulov, G., Mitsuhashi, N.V., Mobarry, C.,
Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L.,
Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K.,
Nusskern, D.R., Pacle, J.M., Palazzolo, M., Pittman, G.S., Pan, S.,
Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K.,
Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I.,
Stapleton, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C.,
Stapleton, M., Strong, R., Sun, E., Svirska, R., Tector, C., Turner, R.,
Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wasserman, D.A.,
Weinstock, G.M., Weissbach, J., Williams, S.M., Woodage, T.,
Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F.,
Zhong, F.J., Zhong, W., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H.,
Zhang, F.N., Myers, E.W., Rubin, G.M., and Venter, J.C.
The genome sequence of *Drosophila melanogaster*
Science 287 (5461), 2185-2195 (2000)
20196006
2 (bases 1 to 298469)
ADAMS, M.D., CELINKER, S.E., GIBBS, R.A., RUBIN, G.M. and VENTER, C.J.
Direct Submission
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
On Oct 9, 2000 this sequence version replaced gi:7304344.
Location/Qualifiers
1..298469
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="4"
complement(1028..>1306)
/gene="CG17469"
/product="CT38623"
/db_xref="FLYBASE:FBan0017469"
/db_xref="FLYBASE:FBgn0040845"
complement(<1028..>1306)
/gene="CG17469"
/map="102E1-102E3"
/db_xref="FLYBASE:FBan0017469"
/db_xref="FLYBASE:FBgn0040845"
complement(1028..1192)
/gene="CG17469"
/note="CG17469 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0017469"
/db_xref="FLYBASE:FBgn0040845"
/protein_id="AAF59375.1"
/db_xref="GI:7304345"
/translation="MTESGIDLCFDMEDLNLINLNDNDNDPLPNTENNPEYELKS
SSRCIRHNEW"
join(3595..3608,3730..3887,3945..4042,4116..4365,
4417..4547,4611..4779,6348..6499,7372..7501,7555..>7671)
/gene="CG17471"
/product="CT38625"
/db_xref="FLYBASE:FBan0017471"
/db_xref="FLYBASE:FBgn0039924"

gene
CDS
3595..>7671
/gene="CG17471"
/map="102E1-102E3"
/db_xref="FLYBASE:FBan0017471"
/db_xref="FLYBASE:FBgn0039924"
join(3750..3887,3945..4042,4116..4365,4417..4547,
4611..4779,6348..6499,7372..7501,7555..7671)
/gene="CG17471"
/note="CG17471 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0017471"
/db_xref="FLYBASE:FBgn0039924"
/protein_id="AAF59376.2"
/db_xref="GI:10728139"
/translation="MEKISSSOPRLKKKHPYKHOKVLFPAENPLISVFMGCTN
HTNELSGVNPVWLPPDPRAYSKIKVDNLFPKKNPSFKYKCYPLFNRNRER
FGVDVYRRESLTISOPLOIDSSGSGAQFQSDTKFTITSLTSEETERNHAPLKQY
HPYVERGRKTLTPQYLGMRYITVESGVYFVKNRNFSSHLTHKKRDLGTVYRE
ASEKELEKNLPTFKDNDIFIKOKVLDIGKEKDKLMDPLTSLNDVLTLEHIMDSLV
GVHDCVRAEEALOGDNLITVGRSENSEECDSGSEFATVTGKIKELMYIINCLIA
EKRELYFALIDVLTQYGVKKQAKAATVYGSNVDSICDPEQVAKRFLDPMKRA
IE"
complement(join(7685..7850,7901..7982,8045..8187,
9658..10140,10199..10336))
/gene="CG17467"
/product="CT38621"
/db_xref="FLYBASE:FBan0017467"
/db_xref="FLYBASE:FBgn0022341"
complement(<7685..>10336)
/gene="CG17467"
/map="102E1-102E3"
/db_xref="FLYBASE:FBan0017467"
/db_xref="FLYBASE:FBgn0022341"
complement(join(7848..7850,7901..7982,8045..8187,
9658..10140,10199..10297))
/gene="CG17467"
/note="CG17467 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0017467"
/db_xref="FLYBASE:FBgn0022341"
/protein_id="AAF59377.1"
/db_xref="GI:7304347"
/translation="MKLDANIKSGLSNHVQYTLDPKKYQIHNQVQENSDEIVKSDS
KPEEVKATIELSKKFEFFYSPEAKNFIKFTMPPEGVVFPPEPSEVNOOIST
LEFNILIOKTKTSTILNKPRMEKQKDOOKKPNPLCTFPPISOPFISDT
LEESDSYQXENSEDESEINPSNYDKALKLAQSAKAKOLRAKTEKQNNIEIO
EIKGRIDVYSQLSNESIESAKAIRERFERMKRSIEMENPSQTQIKRFY"
join(18584..18839,18906..19163,25083..25226,25280..26049,
26111..26402,27005..27124,27178..27740)
/gene="pho"
/note="Nucleotide sequence of the Celera sequence differs
from the published sequence for this transcript."
/product="CT39329"
/db_xref="FLYBASE:FBan0017743"
/db_xref="FLYBASE:FBgn0002521"
complement(<27740..>27740)
/gene="pho"
/note="CG17743"
/map="102E3-102E5"
/db_xref="FLYBASE:FBan0017743"
/db_xref="FLYBASE:FBgn0002521"
join(19069..19163,25083..25226,25280..26049,26111..26402,
27005..27124,27178..27325)
/gene="pho"
/note="pho gene product; Nucleotide sequence of the Celera
sequence differs from the published sequence for this
transcript"
/codon_start=1
/db_xref="FLYBASE:FBan0017743"
/db_xref="FLYBASE:FBgn0002521"
/protein_id="AAF59378.1"
/db_xref="GI:7304348"
/translation="MAYERFGIILIQSEYDEDIGNTRVKNQKMGKGNHYDLHRKNAP

CD5

gene

mRNA

complement{join(33838. .34674,34722. .>35332)}

/gene="CG17461"

/product="CT38613"

/db_xref="FLYBASE:FBan0017461"

/db_xref="FLYBASE:FBgn0039925"

complement{<33838. .>35332}

/gene="CG17461"

/map="102E3-102E5"

/db_xref="FLYBASE:FBan0017461"

/db_xref="FLYBASE:FBgn0039925"

complement{join(33838. .34674,34722. .35255)}

/note="CG17461 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0017461"
 /db_xref="FLYBASE:FBgn0039925"
 /protein_id="AAF59380.1"
 /db_xref="GI:7304350"
 /translation="WTTNVRVRLALVTYLETINERITDRLKNENNTNVIINHPLEKLPET
 GVSPTLTQPVANANQCYDMLHFGNKRKRVTAATLMMKNSSRSHTFTITLESQPELINA
 SIGSAFEGIGIRGKSLDYDLADLAGESEKORTQAGORLKEASQINLSLGLAVSLVAD
 GKAKIVPRDSKLTFLQYDFNPKSLGGNTTATLWCSIPDHYDEITSLRVAASRA
 KNISKKPINDEPKDARLROYQNELLYIKRLMLQSSQIINKNNPNKTIKSPLEKTIQ
 TNMNSTKVNQIITLGDRCNKAASFKTNNSLTILPNPFLQCEEVQLQASRTDLIKRSL
 IGGERTIHQPELKEKHMARRVAAORHLISALALRVYCCEDSDLLQGHVATTTQDIDIKR
 NDYIKKCKEAKIKMLEMEVSDLNSEFQDREDYLDIEIRMLGRQVAFHQQLFLKFSNPGC
 KEDRW"
 complement({join(50956..51220,52591..52741)})

[illegible]

RESULT 8
AC014617

LOCUS	AC014617	21595 bp	DNA	HTG	16-NOV-1999
DEFINITION	<i>Drosophila melanogaster</i> , *** SEQUENCING IN PROGRESS ***, in ordered pieces.				
ACCESSION	AC014617				
VERSION	AC014617.1	GI:6436718			
KEYWORDS	HTG; HTGS-PHASE2.				
SOURCE	fruit fly.				
ORGANISM	<i>Drosophila melanogaster</i>				
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephyridroidea; Drosophilidae; Drosophila.				
AUTHORS	1 (bases 1 to 21595)				
TITLE	Adams,M. and Venter,J.C.				
JOURNAL	Direct Submission				
COMMENT	Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA This sequence was identified as CDM:10212914 by the submitter. For further information on this sequence e-mail to fly@celera.com. * NOTE: This is a 'working draft' sequence. * This sequence will be replaced * by the finished sequence as soon as it is available and * the accession number will be preserved.				

FEATURES	source
Location/Qualifiers	1. .21595
/organism="Drosophila melanogaster"	
/db_xref="taxon:7227"	
BASE COUNT	7103 a 3538 c 3704 g 7250 t
ORIGIN	

Query Match	8.2%;	Score 45.6;	DB 64;	Length 21595;
Best Local Similarity	46.0%;	Pred. No. 6.1;		
Matches 193; Conservative	0;	Mismatches 224;	Indels 3;	Gaps 1;

[illegible]

RESULT	9
AL513530/c	
LOCUS	AL513530 193742 bp DNA HTG 10-FEB-2001
DEFINITION	Homo sapiens chromosome 6 clone RP11-447M22, *** SEQUENCING IN
PROGRESS	***, 16 unordered pieces.
ACCESSION	AL513530
VERSION	AL513530.2 GI:12750914
KEYWORDS	HTG; HTGS_PHASE1.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE
AUTHORS
JOURNAL

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 193742)

TITLE
Direct Submission

Submitted (09-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
On Feb 12, 2001 this sequence version replaced gl:12743866.

COMMENT

Genome Center

Center: Sanger Centre

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk

Project Information

Center project name: ba447M22

Summary Statistics

Assembly program: XGAP; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 189015 bases at least Q40
Consensus quality: 191018 bases at least Q30
Consensus quality: 191870 bases at least Q20
Insert size: 192242; sum-of-contigs
Insert size: 175186; 1.1% error; agarose-fp
Quality coverage: 5.03x in Q20 bases; sum-of-contigs Quality
coverage: 5.74x in Q20 bases; agarose-fp

NOTE: This is a 'working draft' sequence. It currently
consists of 16 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 5494: contig of 5494 bp in length
5495 5594: gap of 100 bp
5595 8469: contig of 2875 bp in length
8470 8569: gap of 100 bp
8570 22023: contig of 13454 bp in length
22024 22123: gap of 100 bp
22124 27549: contig of 5426 bp in length
27550 27649: gap of 100 bp
27650 35659: contig of 8010 bp in length
35660 35759: gap of 100 bp
35760 44155: contig of 8396 bp in length
44156 44255: gap of 100 bp
44256 48345: contig of 4090 bp in length
48346 48445: gap of 100 bp
48446 88503: contig of 40058 bp in length
88504 88603: gap of 100 bp
88604 94516: contig of 5913 bp in length
94517 94616: gap of 100 bp
94617 123775: contig of 29159 bp in length
123776 123875: gap of 100 bp
123876 129937: contig of 6062 bp in length
129938 130037: gap of 100 bp
130038 140301: contig of 10264 bp in length
140302 140401: gap of 100 bp
140402 149963: contig of 9562 bp in length
149964 150063: gap of 100 bp
150064 171500: contig of 21437 bp in length
171501 171600: gap of 100 bp
171601 189885: contig of 18285 bp in length
189886 189985: gap of 100 bp
189986 193742: contig of 3757 bp in length.

FEATURES

Source

1. 193742

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="6"

/clone="RP11-447M22"

misc-feature /clone_id="RPC1-11.2"
1. 5494
/note="assembly_fragment:00732
fragment_chain:1"
misc-feature 5595. 8469
/note="assembly_fragment:01506
fragment_chain:1"
misc-feature 8570. 22023
/note="assembly_fragment:01113
fragment_chain:1"
misc-feature 22124. 27549
/note="assembly_fragment:01623
fragment_chain:1"
misc-feature 27650. 35659
/note="assembly_fragment:01123
fragment_chain:1"
misc-feature 35760. 44155
/note="assembly_fragment:01583
fragment_chain:2"
misc-feature 44256. 48345
/note="assembly_fragment:01531
fragment_chain:2"
misc-feature 48446. 88503
/note="assembly_fragment:00536
fragment_chain:2"
misc-feature 88604. 94516
/note="assembly_fragment:00318
fragment_chain:2"
misc-feature 94617. 123775
/note="assembly_fragment:01777
fragment_chain:2"
misc-feature 123876. 129937
/note="assembly_fragment:00240
fragment_chain:3"
misc-feature 130038. 140301
/note="assembly_fragment:01550
fragment_chain:3"
misc-feature 140402. 149963
/note="assembly_fragment:00827
fragment_chain:3"
misc-feature 150064. 171500
/note="assembly_fragment:00478
fragment_chain:3"
misc-feature 171601. 189885
/note="assembly_fragment:00881"
189986. 193742
/note="assembly_fragment:01274"
BASE COUNT 62033 a 34704 c 32982 g 62522 t 1501 others
ORIGIN
Query Match 8.2% Score 45.6: DB 82: Length 193742;
Best Local Similarity 46.3% Pred: No. 5;
Matches 252; Conservative 0; Mismatches 289; Indels 3; Gaps 3;
QY 1 ATGCTTTAAACACACCTGATGATATATATATCTCAGCTTTCTTC 60
DB 88131 ATATATTTATATATATATATATATATATATATATATATATAT 88072
QY 61 CCATTAAACATGATGATATGCTATGCTATGCTATGCTATGCTATG 120
DB 88071 ATATATTTTACATATGCTTTATATATATATATATATATATAT 88012
QY 121 GCATTAAAGCATGAGCATGCTGTTAAACAGCTGCCAATATTCATC 180
DB 88011 ATATATTTATATATATATATATATATATATATATATATATAT 87953
QY 181 CTTCGAATGCTGCTGATATATGTTATATGTTTAAAGCTTATTCG 240
DB 87952 TTTATATATGTTTATATATATATATATATATATATATATATAT 87893
QY 241 AGAGTTAAAGATCTTTAAATTCATATCCATGAGGTGATTTCTTAA 299
CTTAATATAGGGGTTTT

* 9419 9518: gap of 100 bp
* 9519 10521: contig of 1003 bp in length
* 10522 10621: gap of 100 bp
* 10622 11585: contig of 964 bp in length
* 11586 11685: gap of 100 bp
* 11686 12674: contig of 989 bp in length
* 12675 12774: gap of 100 bp
* 12775 13775: contig of 1001 bp in length
* 13776 13875: gap of 100 bp
* 13876 14783: contig of 908 bp in length
* 14784 14883: gap of 100 bp
* 14884 15865: contig of 982 bp in length
* 15866 15965: gap of 100 bp
* 15966 16944: contig of 979 bp in length
* 16945 17044: gap of 100 bp
* 17045 17985: contig of 941 bp in length
* 17986 18085: gap of 100 bp
* 18086 19048: contig of 963 bp in length
* 19049 19148: gap of 100 bp
* 19149 20106: contig of 958 bp in length
* 20107 20206: gap of 100 bp
* 20207 21176: contig of 970 bp in length
* 21177 21276: gap of 100 bp
* 21277 22256: contig of 980 bp in length
* 22257 22356: gap of 100 bp
* 22357 23336: contig of 980 bp in length
* 23337 23436: gap of 100 bp
* 23437 24405: contig of 969 bp in length
* 24406 24505: gap of 100 bp
* 24506 25457: contig of 952 bp in length
* 25458 25557: gap of 100 bp
* 25558 26541: contig of 964 bp in length
* 26542 26641: gap of 100 bp
* 26642 27615: contig of 974 bp in length
* 27616 27715: gap of 100 bp
* 27716 28695: contig of 980 bp in length
* 28696 28795: gap of 100 bp
* 28796 29720: contig of 925 bp in length
* 29721 29820: gap of 100 bp
* 29821 30786: contig of 966 bp in length
* 30787 30886: gap of 100 bp
* 30887 31838: contig of 952 bp in length
* 31839 31938: gap of 100 bp
* 31939 32885: contig of 947 bp in length
* 32886 32985: gap of 100 bp
* 32986 33960: contig of 975 bp in length
* 33961 34060: gap of 100 bp
* 34061 35062: contig of 1002 bp in length
* 35063 35162: gap of 100 bp
* 35163 36155: contig of 997 bp in length
* 36160 36259: gap of 100 bp
* 36260 37224: contig of 965 bp in length
* 37225 37324: gap of 100 bp
* 37325 38293: contig of 969 bp in length
* 38294 38393: gap of 100 bp
* 38394 39377: contig of 984 bp in length
* 39378 39477: gap of 100 bp
* 39478 40410: contig of 933 bp in length
* 40411 40510: gap of 100 bp
* 40511 41478: contig of 968 bp in length
* 41479 41578: gap of 100 bp
* 41579 42553: contig of 975 bp in length
* 42554 42653: gap of 100 bp
* 42654 43612: contig of 959 bp in length
* 43613 43712: gap of 100 bp
* 43713 44689: contig of 977 bp in length
* 44690 44789: gap of 100 bp
* 44790 45718: contig of 929 bp in length
* 45719 45818: gap of 100 bp
* 45819 46747: contig of 929 bp in length
* 46748 46847: gap of 100 bp
* 46848 47815: contig of 968 bp in length
* 47816 47915: gap of 100 bp

* 47916 48930: contig of 1015 bp in length
* 48931 49030: gap of 100 bp
* 49031 50005: contig of 975 bp in length
* 50006 50105: gap of 100 bp
* 50106 51093: contig of 998 bp in length
* 51094 51193: gap of 100 bp
* 51194 52181: contig of 988 bp in length
* 52182 52281: gap of 100 bp
* 52282 53272: contig of 991 bp in length
* 53273 53372: gap of 100 bp
* 53373 54370: contig of 998 bp in length
* 54371 54470: gap of 100 bp
* 54471 55469: contig of 999 bp in length
* 55470 55569: gap of 100 bp
* 55570 56544: contig of 975 bp in length
* 56545 56644: gap of 100 bp
* 56645 57668: contig of 1024 bp in length
* 57669 57768: gap of 100 bp
* 57769 58728: contig of 960 bp in length
* 58729 58828: gap of 100 bp
* 58829 59829: contig of 1001 bp in length
* 59830 59929: gap of 100 bp
* 59930 60913: contig of 984 bp in length
* 60914 61013: gap of 100 bp
* 61014 62013: contig of 1000 bp in length
* 62014 62113: gap of 100 bp
* 62114 63077: contig of 964 bp in length
* 63078 63177: gap of 100 bp
* 63178 64116: contig of 939 bp in length
* 64117 64216: gap of 100 bp
* 64217 65161: contig of 945 bp in length
* 65162 65261: gap of 100 bp
* 65262 66233: contig of 972 bp in length
* 66234 66333: gap of 100 bp
* 66334 67313: contig of 980 bp in length
* 67314 67413: gap of 100 bp
* 67414 68394: contig of 981 bp in length
* 68395 68494: gap of 100 bp
* 68495 69484: contig of 990 bp in length
* 69485 69584: gap of 100 bp
* 69585 70506: contig of 922 bp in length
* 70507 70606: gap of 100 bp
* 70607 71547: contig of 941 bp in length
* 71548 71647: gap of 100 bp
* 71648 72629: contig of 982 bp in length
* 72630 72729: gap of 100 bp
* 72730 73654: contig of 925 bp in length
* 73655 73754: gap of 100 bp
* 73755 74680: contig of 926 bp in length
* 74681 74780: gap of 100 bp
* 74781 75782: contig of 1002 bp in length
* 75783 75882: gap of 100 bp

Query Match 8.18; Score 44.8; DB 67; Length 84346;
Best Local Similarity 46.28; Pred. No. 7.9;
Matches 148; Conservative 0; Mismatches 172; Indels 0; Gaps 0;

OY 110 TATGGAATATGCAATTAAGCATGATGACATGCGTTAAAGACAGCTGCTCCCAATAT 169
Db 61163 TGTATAGCATTCCTCTTCTCTGTAAGCTTGCAGTTATATGTTTATGACTATTT 61222
OY 170 TCAATCTAGCTCTTGCATTTGGTGGATACATTAATGTTATGTTGTTTAAAGAAATCTT 229
Db 61223 AAAAATATGATCCGACCTGTGTGATGTATATTCATATGTTGTTTGAATTAATTT 61282
OY 230 ATTGCTGAGTAGAGTTAAAGATCTTTAATCATATCCATGGGTATTTCTTAAGTTAA 289
Db 61283 CTTGATGTTTATGATGATATGATGATGATGATGATGATGATGATGATGATGATGAT 61342
OY 290 TAGGGTTTTCAGAGAGATATATGTTGATGATATATTCCTGATGATGATGATGATGAT 349
Db 61343 TTGAGAGATATTCCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 61402

```

OY      350 TATCATATATACATATTTCATATCATACTTGACAAATCATGATGCTGTTTC 409
          || || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      61403 CATTCATTATTTTAACTGATTCGTGATATGATTCGTCGATCATGATGTT 61462
          || || | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      410 TACTATATATAGTCATATT 429
          || || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      61463 TGCACAATATTTTCTCCATT 61482
          || || | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 13
HASMT/c 27694 bp DNA circular PLN 27-OCT-2000
LOCUS    Pichia canadensis mitochondrial DNA, complete sequence.
DEFINITION
D31785
VERSION  D31785.1 GI:1000971
KEYWORDS  ATPase subunit 6; ATPase subunit 8; ATPase subunit 9; NADH
          dehydrogenase subunit 1; NADH dehydrogenase subunit 2; NADH
          dehydrogenase subunit 3; NADH dehydrogenase subunit 4; NADH
          dehydrogenase subunit 4L; NADH dehydrogenase subunit 5; NADH
          dehydrogenase subunit 6; apocytochrome b; cytochrome c oxidase
          subunit I; cytochrome c oxidase subunit II; cytochrome c oxidase
          subunit III; endonuclease homologue protein; large subunit
          ribosomal RNA; small subunit ribosomal RNA; tRNA-Ala; tRNA-Asn;
          tRNA-Asp; tRNA-Cys; tRNA-Gln; tRNA-Glu; tRNA-Gly; tRNA-His;
          tRNA-Ile; tRNA-Leu2; tRNA-Lys; tRNA-Phe; tRNA-Pro; tRNA-Thr;
          tRNA-Trp; tRNA-Tyr; tRNA-Val; transfer RNA-Arg-1; transfer
          RNA-Arg-2; transfer RNA-Leu1; transfer RNA-Met3; transfer RNA-Ser1;
          transfer RNA-Ser2.
SOURCE    Pichia canadensis (strain:21) mitochondrion DNA.
ORGANISM  Mitochondrion Pichia canadensis
          Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
          Saccharomycetaceae; Pichia.
REFERENCE
AUTHORS   Yoshida,K.
TITLE      The structure and organization of mitochondrial genome in Hansenula
          wingel
JOURNAL    International Botanical Congress, 503-504 (1993)
REFERENCE
AUTHORS   Sekito,T., Okamoto,K., Kitano,H. and Yoshida,K.
TITLE      Yeast Hansenula wingel mitochondrial genome's complete DNA sequence
          demonstrated unique characteristics
JOURNAL    Nucleic Acids Symp. Ser. 31, 233-234 (1994)
REFERENCE
AUTHORS   Sekito,T., Okamoto,K., Kitano,H. and Yoshida,K.
TITLE      3 (bases 1 to 27694)
          The complete mitochondrial DNA sequence of Hansenula wingel reveals
          new characteristics of yeast mitochondria
JOURNAL    Curr. Genet. 28 (1), 39-53 (1995)
REFERENCE
AUTHORS   Yoshida,K.
TITLE      4 (sites)
          96022424
JOURNAL    MEDLINE
AUTHORS   Sekito,T., Okamoto,K., Kitano,H. and Yoshida,K.
TITLE      Clone bank of the mitochondrial genome of yeast Hansenula wingel
JOURNAL    Yeast 11 (13), 1317-1321 (1995)
REFERENCE
AUTHORS   Yoshida,K.
TITLE      5 (sites)
          Okamoto,K., Sekito,T. and Yoshida,K.
          The secondary structure and phylogenetic relationship deduced from
          complete nucleotide sequence of mitochondrial small subunit rRNA in
          yeast Hansenula wingel
JOURNAL    Genes Genet. Syst. (1996) In press
JOURNAL    Direct Submission
          Submitted (07-JUN-1994) to the DDBJ/EMBL/GenBank databases. Kazuo
          Yoshida, Hiroshima University, Department of Biological Science;
          1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739, Japan
          (E-mail:K31582@PNIHRO.BINET, Tel:0824-24-7455, Fax:0824-24-0734)
FEATURES
source    1. 27694
           /organism="Pichia canadensis"
           /organelle="mitochondrion"
           /strain="21"
           /db_xref="taxon:4907"
           /note="synonym:Hansenula wingel"

```

```

exon     1..502
         /gene="LSU"
         /number=2
         547..619
         /gene="tRNA-T"
         /note="anticodon-ugu"
         /evidence=not_experimental
         /product="transfer RNA-Thr"
         547..619
         /gene="tRNA-T"
         714..785
         /gene="tRNA-E"
         /note="anticodon-uuc"
         /evidence=not_experimental
         /product="transfer RNA-Glu"
         714..785
         /gene="tRNA-E"
         832..903
         /note="anticodon-gca"
         /evidence=not_experimental
         /product="transfer RNA-Cys"
         832..903
         /gene="tRNA-C"
         1241..3847
         /gene="COX1"
         1241..1963
         /gene="COX1"
         /number=1
         /evidence=not_experimental
         join(1241..1963,2963..3847)
         /gene="COX1"
         /transl_table=1
         /evidence=not_experimental
         /product="cytochrome c oxidase subunit 1"
         /protein_id="BAA06563.2"
         /db_xref="GI:11036337"
         /translation="MYIORWLYSTNAKDIALIYFIFATIFSGVIGSTMSLIIRLEIAP
         GNOLIHGHOLFNVLYVGHALLMIFLWMPGLVGFGRYVMPLILGADMSARLNTI
         SPWLIPPAVLVCVASTLTIESNAGTGMWTPPLUSGIOAHSSPSVDLGIATLHITISL
         LGALNFATSYNMRNGMSYSKMPLEWVAIIITAVMLLSPLVLAGVTMLMDNFN
         TSFEVAGGDPVLVYOHLFWFGHPEVYLIYVPGIASHIVSYSKPVGELSMY
         AMASIAFGFVLVSHMYIVGADTRAYFTSSTWYAVPTGIRKJFSMLATLYGSIR
         LAVPMLVIAFLFLFTIGLGVALANSLDVAFDYVYVGHFHYVLSMGAIFSLA
         GYVNSPOLIGYFENERAQIQFMIIPYGANVTPEMPFLGIQGRPRIPDPDYAG
         WYVSSIGSVTAIISLALFIYITDQLINGLTKIDNKSIVYSKAPDFVESVTFPANN
         SIKSASIEFLNSPPAIHSFNPVAVS"
         <1964..2923
         /gene="COX1"
         /note="no ATG start codon"
         /codon_start=1
         /transl_table=4
         /evidence=not_experimental
         /product="Intronic ORP"
         /protein_id="BAA06564.2"
         /db_xref="GI:11036338"
         /translation="LYIMPWYEIINLYPAICWKLIIIKKFKLITTIYISLYNNIL
         IKLALIVKILYIYNNQOITNNMIKIRKYSKLVGISETIRKNSDLKDKIKFQWML
         AGLIDGVLGVSKAGYTSCEITVALEDEKALROIQNFGGSIKIRSGIKAIRYLHN
         KEGMIKLINAINGINIRNTRKLPOLHKVCSILNLPITPEINLTINNSMFIQFADAGTI
         NYVFKNNYPOLTISVSNKKYIDLKPFMDIFGNIYIDKSNNGCYKWSIQSKTDILFI
         NSYVFNSSRTKRFKMLKCSKYVYLINIKAMLIINKKFNSE"
         1964..2962
         /gene="COX1"
         /cons_splice=(5'site:no,3'site:no)
         /number=1
         /evidence=not_experimental
         2963..3847
         /gene="COX1"
         /number=2
         /evidence=not_experimental
         4115..4261

```

[illegible]

REFERENCE 1 (bases 1 to 118313)
 AUTHORS Murray,J., Rohlfing,T. and Antoniou,B.
 TITLE The sequence of H. sapiens BAC clone CTB-167B5
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 118313)
 AUTHORS Waterston,R.
 TITLE Direct Submission
 JOURNAL Submitted (14-JAN-1998) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 REFERENCE 3 (bases 1 to 118313)
 AUTHORS Waterston,R.
 TITLE Direct Submission
 JOURNAL Submitted (03-FEB-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 REFERENCE 4 (bases 1 to 118313)
 AUTHORS Waterston,R.
 TITLE Direct Submission
 JOURNAL Submitted (04-FEB-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 COMMENT ----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc
 Contact: sapiens@watson.wustl.edu
 Summary Statistics
 Center project name: H_RG167B05

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
 The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see
<http://www.nhgri.nih.gov/DIR/CTB/CHR7> or send
<mailto:egreen@nhgri.nih.gov>

SOURCE INFORMATION:
 Clone CTB-167B5 is from the first release of the human BAC library CTB-978SK-B. The library contains cloned DNA from the male fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U.-J. Kim et al., Genomics 34:213-8 (1996). This clone is available from Research Genetics, Inc. (<http://www.resgen.com>).
 VECTOR: pBelorAC11
 Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:
 The actual start of this clone is at base position 1 of CTB-167B5; actual end of this clone is at base position 118313 of CTB-167B5.

This clone contains SRS SWS52908 (NID:91113638).

FEATURES

source

Location/Qualifiers
 1. 118313
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="7"
 /clone="CTB-167B5"
 /clone_1lb="CTB-978SK-B"
 /map="7q21"

```

repeat_region      353..519
/rpt_family="MER1_type"
repeat_region      527..626
/rpt_family="MER1_type"
repeat_region      686..832
/rpt_family="MER1_type"
repeat_region      914..999
/rpt_family="L2"
repeat_region      1073..1117
/rpt_family="L2"
repeat_region      1359..1503
/rpt_family="L2"
repeat_region      3917..4205
/rpt_family="Alu"
repeat_region      4701..4865
/rpt_family="Alu"
gene               complement(<5050..7954)
CDS                /gene="SRI"
                   complement(join(<5050..5119,6787..6870,7904..7954))
                   /gene="SRI"
                   /note="match to sorcin protein P30626 (PID:g267021);
                   H_RG167B05.1"
                   /codon_start=1
                   /product="calcium binding protein amplified in
                   multidrug-resistant cells"
                   /protein_id="AAB96327.1"
                   /db_xref="GI:2772536"
                   /translation="MAYPGHAGAGGCGYPCGYGAGPAGPAGPAGTODPLXGFAAVAG
                   ODQIDADEORCLTOSGINGYK"
repeat_region      9625..10516
/rpt_family="L1"
repeat_region      10518..11038
/rpt_family="L1"
repeat_region      11200..11288
/rpt_family="MIR"
repeat_region      11353..11404
/rpt_family="MER1_type"
repeat_region      11466..11711
/rpt_family="MER1_type"
repeat_region      12803..12913
/rpt_family="MER1_type"
repeat_region      13488..13527
/rpt_family="MIR"
repeat_region      15273..15566
/rpt_family="Alu"
repeat_region      15767..15829
/rpt_family="MER1_type"
repeat_region      15947..16238
/rpt_family="Alu"
repeat_region      16244..16326
/rpt_family="MER1_type?"
repeat_region      16327..16913
/rpt_family="L1"
repeat_region      16912..17147
/rpt_family="L1"
repeat_region      17199..17400
/rpt_family="MaLR"
repeat_region      17401..17517
/rpt_family="MER1_type?"
repeat_region      18465..18536
/rpt_family="L1"
repeat_region      18644..18689
/rpt_family="(CA)n"
repeat_region      19101..19157
/rpt_family="MIR"
repeat_region      19436..20346
/rpt_family="L1"
repeat_region      20347..20650
/rpt_family="Alu"
repeat_region      21158..21226
/rpt_family="L1"
repeat_region      21425..21738
/rpt_family="Alu"

```

Job time: 4726 sec

```
repeat_region 21816..21903  
/rpl_family="MER81"  
repeat_region 21910..22079  
/rpl_family="MER1_type"  
repeat_region 22715..23122  
/rpl_family="L1"  
repeat_region 24950..25239  
/rpl_family="Alu"  
repeat_region 25318..25618  
/rpl_family="Alu"  
repeat_region 25673..25862  
/rpl_family="L1"  
misc_feature 26008..26260  
/note="match to EST AA327226 (MID:g1979532)"  
repeat_region 26261..26341  
/rpl_family="MalR"  
misc_feature 26342..26414  
/note="match to EST AA327226 (MID:g1979532)"  
repeat_region 26564..26742  
/rpl_family="L1"  
repeat_region 26853..27102  
/rpl_family="L1"  
repeat_region 27108..27407  
/rpl_family="Alu"  
repeat_region 28496..28701  
/rpl_family="Alu"  
repeat_region 28961..29023  
/rpl_family="(TGA)n"  
repeat_region 29217..29523  
/rpl_family="Alu"  
repeat_region 30548..30650  
/rpl_family="MIR"  
repeat_region 32058..32121  
/rpl_family="L1"  
repeat_region 32134..32276  
/rpl_family="L1"  
repeat_region 33240..36502  
/rpl_family="L1"  
repeat_region 36503..36633  
/rpl_family="Alu"  
repeat_region 36634..37363  
/rpl_family="L1"  
repeat_region 37394..37548  
/rpl_family="MalR"  
repeat_region 37657..38131
```

Query Match 8.0%: Score 44.2; DB 85; Length 118313;
Best Local Similarity 49.0%: Pred.No.10; Mismatches 123; Indels 0; Gaps 0;

```
Matches 118; Conservative 0; Mismatches 123; Indels 0; Gaps 0;  
  
QY 32 TGAATATATTTCACCTTTCTTCCATTAACATGATCGCATTAAGTATCTCTA 91  
|| ||| ||||| || ||| || |||| |||| |||| |||| |||| ||||  
DB 100718 TCGTAAAGTTATTTCTTTTGGGTTCTTTAAAGATATATATATTTTTR 100777  
  
QY 92 TGTGTCATGTTTAAATATGGAATATGCAATTAAGCATGATGCTGTTAAA 151  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 100778 TACTGGTACTATTAAGTATGATCTGTAGTACCAATACAGATGATTTATGTTTAC 100837  
  
QY 152 GCAGTGTCTCCATATCATCTAGGTCCTGCAATGCTGGGATACATAATGTTATAG 211  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 100838 AGAGTGTGGCTAAACAAATTTATAGTACGTAGATGCTTTACATATATTAACACTT 100897  
  
QY 212 TTGATTAGAAATCTTTTTCGTGGAGTAGAGTTAAAGATCTTAATCATATCCATG 271  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 100898 TAGTTAAATTAATTAACATAATATTAAACAATTATGCAAAATTCAGATATAATG 100957  
  
QY 272 G 272  
DB 100958 G 100958
```

Search completed: May 30, 2001, 23:07:52

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 30, 2001, 22:23:47 ; Search time 113.14 Seconds

(without alignments)
2863.697 Million cell updates/sec

Title: US-09-382-242-23

Perfect score: 555
Sequence: 1 ATGTCCTTAACAAGCAGCTC.....AATCATACACTAAACGATAG 555

Scoring table: IDENTITY NUC
Gap 10.0 , Gapext 1.0

Searched: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

_N_Geneseq_0401.*
1: /cgnl_8/gcgdata/geneseq/geneseqn/NA1980.DAT.*
2: /cgnl_8/gcgdata/geneseq/geneseqn/NA1981.DAT.*
3: /cgnl_8/gcgdata/geneseq/geneseqn/NA1982.DAT.*
4: /cgnl_8/gcgdata/geneseq/geneseqn/NA1983.DAT.*
5: /cgnl_8/gcgdata/geneseq/geneseqn/NA1984.DAT.*
6: /cgnl_8/gcgdata/geneseq/geneseqn/NA1985.DAT.*
7: /cgnl_8/gcgdata/geneseq/geneseqn/NA1986.DAT.*
8: /cgnl_8/gcgdata/geneseq/geneseqn/NA1987.DAT.*
9: /cgnl_8/gcgdata/geneseq/geneseqn/NA1988.DAT.*
10: /cgnl_8/gcgdata/geneseq/geneseqn/NA1989.DAT.*
11: /cgnl_8/gcgdata/geneseq/geneseqn/NA1990.DAT.*
12: /cgnl_8/gcgdata/geneseq/geneseqn/NA1991.DAT.*
13: /cgnl_8/gcgdata/geneseq/geneseqn/NA1992.DAT.*
14: /cgnl_8/gcgdata/geneseq/geneseqn/NA1993.DAT.*
15: /cgnl_8/gcgdata/geneseq/geneseqn/NA1994.DAT.*
16: /cgnl_8/gcgdata/geneseq/geneseqn/NA1995.DAT.*
17: /cgnl_8/gcgdata/geneseq/geneseqn/NA1996.DAT.*
18: /cgnl_8/gcgdata/geneseq/geneseqn/NA1997.DAT.*
19: /cgnl_8/gcgdata/geneseq/geneseqn/NA1998.DAT.*
20: /cgnl_8/gcgdata/geneseq/geneseqn/NA1999.DAT.*
21: /cgnl_8/gcgdata/geneseq/geneseqn/NA2000.DAT.*
22: /cgnl_8/gcgdata/geneseq/geneseqn/NA2001.DAT.*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	555	100.0	555	18 T79321	DNA encoding Staph
2	42.2	7.6	3567	21 A70117	Plasmodium falcipa
3	40.6	7.3	4197	21 A70227	Plasmodium falcipa
4	39.4	7.1	3927	21 A70101	Plasmodium falcipa
5	39.2	7.1	615	20 X99599	Nucleic acid sequ
6	39.2	7.1	5391	19 V71729	Upstream sequence
7	39.2	7.1	5407	19 V71738	Upstream sequence
8	38.4	6.9	5423	18 W74335	Staphylococcus aur
9	38.2	6.9	1612	7 N60392	Sequence encoding
10	38	6.8	8855	20 X02997	Human IL-1ra BAC c
11	38	6.8	2866	20 X22304	Human IL-1ra BAC c

C 12	37.8	6.8	1444	20	Z17016
C 13	37.6	6.8	6042	21	A70199
C 14	36.6	6.6	1536	16	O94111
C 15	36.6	6.6	1864	15	O78892
C 16	36.4	6.6	3413	20	W72330
C 17	36.4	6.6	1082138	21	F22305
C 18	36	6.5	1452	21	A70123
C 19	36	6.5	2854	19	V53366
C 20	36	6.5	4673	13	O27189
C 21	35.8	6.5	5409	21	A70151
C 22	35.8	6.5	910715	20	X20248
C 23	35.8	6.5	1664976	19	V21209
C 24	35.6	6.4	1218	19	V29579
C 25	35.6	6.4	12577	19	V35616
C 26	35.6	6.4	32367	19	V35616
C 27	35.6	6.4	94618	21	F22285
C 28	35.2	6.3	1768	21	C74378
C 29	35	6.3	716	20	X13483
C 30	35	6.3	1111	18	V75004
C 31	35	6.3	7059	15	O44750
C 32	35	6.3	10322	18	V74409
C 33	35	6.3	867	21	A02283
C 34	34.6	6.2	2486	21	A97037
C 35	34.6	6.2	4968	21	Z55696
C 36	34.6	6.2	111309	20	X20250
C 37	34.6	6.2	111309	20	X20250
C 38	34.6	6.2	111309	20	X20248
C 39	34.6	6.2	910715	20	X20248
C 40	34.4	6.2	881	20	Z16739
C 41	34.4	6.2	5327	19	T99971
C 42	34.4	6.2	5400	21	A39732
C 43	34.4	6.2	9789	17	T41852
C 44	34.4	6.2	22960	20	X13282
C 45	34.2	6.2	6033	21	A70152

ALIGNMENTS

RESULT 1	T79321	standard; DNA; 555 BP.
XX	XX	
AC	T79321:	
XX	XX	
DT	16-FEB-1998	(first entry)
XX	XX	
DE	DNA encoding Staphylothermus marinus esterase Fl-12LC.	
XX	XX	
KW	Esterase; thermostable enzyme; ester; chiral compound; cheese;	
KW	pulp; paper; lignin removal; sugar; lignocellulose;	
KW	disease resistance; feedstuff; ss.	
OS	Staphylothermus marinus strain Fl.	
XX	XX	
FT	Key	Location/Qualifiers
FT	CDS	1..576
FT	FT	/*tag= a
XX	XX	/transl_except= (pos: 526..528, aa:ASP)
PN	W09730160-A1.	
PD	21-AUG-1997.	
XX	XX	
PF	11-FEB-1997;	97WO-US02039.
XX	XX	
PR	16-FEB-1996;	96US-0602359.
XX	XX	
PA	(RECO-) RECOMBINANT BIOCATALYSIS INC.	
XX	XX	
PI	Callen W, Kosmicka A, Link S, Maffia AM, Murphy D;	
XX	Reid J, Robertson DE, Swanson RV, Warren PV;	
XX	XX	

Human gene express
Plasmodium falcipa
mmL genomic DNA.
Aspergillus nidula
F. balustinum CP70
Arabidopsis thalia
Plasmodium falcipa
DNA encoding a sta
P. yoelii SSP2 ant
Plasmodium falcipa
Borrelia burgdorfe
Methanococcus jann
pythium oligandrum
SHOX gene prelinin
Human SHOX (short
BAC containing rep
Human secreted pro
Enterococcus faeca
Staphylococcus aur
Borrelia burgdorfe
Soybean lipoxigena
Staphylococcus aur
Human colon cancer
Nucleotide sequenc
Heliothis armigera
Borrelia burgdorfe
Borrelia burgdorfe
Borrelia burgdorfe
Human gene express
Zix gene. Bos tau
Potato subclone pg
cDNA encoding Plas
Enterococcus faeca
Plasmodium falcipa


```
Db 1193 ataataatgaagaagaaattcttatgaaaaataaagtcaaggaggatccca 1252
QY 91 ATGTCGTCATGCTTTAATATATGAAATATGCAATTAAGCAGATCTAGCATTA 150
Db 1253 ttcttaagagataaaaaataatgctglaaataatgagtgaggctataagaatcgaag 1312
QY 151 AGCAGTGTGGTCCCAATATTAATCTAGGCTTGCATTTGGTGGATCTACTATTTTATA 210
Db 1313 atgatatgataagatataatatacaagaagatattacaaggctgaggtatcatatgata 1372
QY 211 GTTGCTTTAAGAAATCTTATTCGTCGAGTAGAGTTAAAGGATCTTTATCATATCCATG 270
Db 1373 atataataataaataatgataataataataataatgataataataataataatg 1432
QY 271 GGTGATTTCTTAACCTTAATAGGGGTTTCGACGAGATATATGTTGGATCATTTCTTA 330
Db 1433 ataataataataataatgataataataataataataataataataataataata 1492
QY 331 GTCACGATTTGTTTCTTTTATCAATATAGCATATTTTCATAGCTATATCAATCTTAC 390
Db 1493 atgataataataataataataataataataataataataataataataataata 1552
QY 391 AATCATGATGAGCTGTCTCTAATATAG 421
Db 1553 ataataataataataataataatctg 1583

RESULT 3
A70227/c
ID A70227 standard; DNA; 4197 BP.
XX
AC A70227;
XX
DT 07-NOV-2000 (first entry)
DE Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:360.
XX
KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KM antimalarial; malaria; protozoacide; infection; insecticide; ds.
XX
OS Plasmodium falciparum.
XX
PN WO200025728-A2.
XX
PD 11-MAY-2000.
XX
PE 05-NOV-1999; 99WO-US26796.
XX
PR 05-NOV-1998; 98US-0107131.
XX
PA (HOFF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX
PI Hoffman S, Carucci D, Gardner M, Venter JC;
XX
DR WPI: 2000-365347/31.
XX
PT Proteins encoded by chromosome 2 of the human malaria parasite,
XX Plasmodium falciparum, useful as antimalarial vaccines and in the
XX diagnosis of P.falciparum infection -
XX
PS Disclosure; Page 548-549; 577pp; English.
XX
CC The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malaria parasite, Plasmodium falciparum.
CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (1) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (1), are
CC useful in the detection of infection with P. falciparum. Furthermore,
```

```
CC (I) (especially when they are rifins or secreted or membrane proteins)
CC can aid the identification of drugs to treat or prevent P. falciparum
CC infection, or they can be used to identify drug resistance in
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC subsequent identification of proteins encoded by it will help to expand
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasite life cycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. A70078 to A70287 and B18144 to B18352 represent nucleotide and
CC protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.
XX
SQ Sequence 4197 BP; 2005 A; 334 C; 525 G; 1333 T; 0 other;

Query Match 7.3%; Score 40.6; DB 21; Length 4197;
Best Local Similarity 46.1%; Pred. No. 0.44;
Matches 136; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

QY 125 TAAGGATCTGAGACATGCTTTAAAGCAGTGTCCCATATTCATCTAGGCTTG 184
Db 1441 TAACAAAATTATGAATATCCATAAAAATATATATCCAGAACGATTTATATTAACATAT 1382
QY 185 CAATTGTGGGATACTAATTTTATAGTTGTTTAAAGAAATCTTTCGTCGACTAGAG 244
Db 1381 CATTTGAAAATATAAATTTGACAGATGGTTCATCATATTTTCTTAATATATTA 1322
QY 245 TTAAGGATCTTTAATCAATATCCATGCGTGATTTCTTAACCTTAATAGGGCTTTTCAGC 304
Db 1321 TTTTATGAGGTTTAAAAAATCGTTAATTAATCATTCGTTTTCCTATAACAA 1262
QY 305 AAGTATATGTTGATACATTCCTAGTCTCAGTATTTGTTTCTTATCAATAATAGCAT 364
Db 1261 GATATTTTTCATGATTAAGATTTTCTTTTCTTTTCTTTTCTTATATCAATATAT 1202
QY 365 ATTTCATGCTATATCAATATCTGACAAATCAATGATGATCTGTTCTACTAATAAT 419
Db 1201 TGGATAATTGATTAGGAGATTTATTGATGATTCCTATATTCTCTCTTTTAT 1147

RESULT 4
A70101/c
ID A70101 standard; DNA; 3927 BP.
XX
AC A70101;
XX
DT 07-NOV-2000 (first entry)
DE Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:234.
XX
KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KM antimalarial; malaria; protozoacide; infection; insecticide; ds.
XX
OS Plasmodium falciparum.
XX
PN WO200025728-A2.
XX
PD 11-MAY-2000.
XX
PE 05-NOV-1999; 99WO-US26796.
XX
PR 05-NOV-1998; 98US-0107131.
XX
PA (HOFF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX
PI Hoffman S, Carucci D, Gardner M, Venter JC;
XX
DR WPI: 2000-365347/31.
```

XX Proteins encoded by chromosome 2 of the human malarial parasite,
PT *Plasmodium falciparum*, useful as antimalarial vaccines and in the
XX diagnosis of *P.falciparum* infection -
XX
PS *Disclosure: Page 458-459; 577pp; English.

Disclosure; page 458-459; 577pp; English.

The present invention describes proteins and their fragments (I) encoded by chromosome 2 of the human malarial parasite, *Plasmodium falciparum*. Also described are: (1) nucleotide sequences (II) encoding (1); and (2) vaccines against *P. falciparum* infection comprising (1) or (II).

(1) and (II) are useful for the development of vaccines against *P. falciparum* infection. (1) and polyclonal antisera or a monoclonal antibody raised to immunogens comprising the sequences of (1), are useful in the detection of infection with *P. falciparum*. Furthermore, (1) (especially when they are rifins or secreted or membrane proteins) can aid the identification of drugs to treat or prevent *P. falciparum* infection, or they can be used to identify drug resistance in *P. falciparum*. Sequencing of the *Plasmodium* chromosome 2 and the subsequent identification of proteins encoded by it will help to expand our understanding of parasite biology, a process hampered by the complexity of the parasitic life-cycle, and provide new targets for vaccine and drug development. Parasite resistance to drugs and mosquito resistance to insecticides have led to a resurgence of malaria in many parts of the world, and there is a pressing need for vaccines and new drugs. A70078 to A70287 and B18144 to B18352 represent nucleotide and protein sequences given in the present invention, but which are not specifically mentioned within the specification.

Sequence 3927 BP; 1786 A; 198 C; 683 G; 1260 T; 0 other;

Query Match	7.1%	Score 39.4	DB 21	Length 3927
Best Local Similarity	44.2%	Pred. No. 0.86		
Matches 163, Conservative	0	Mismatches 206	Indels 0	Gaps 0

OY	52	TTTTCTTCCCATTAACAAGATGCGATTCAGCTATCCCTGATGCGCATAGGGTATATATA	11.1
Db	767	TTATTAATTAATCTTATTAATAATTTCAAAATTAATCGTTAAATTAATATCATCATGATTAAAA	7089
OY	112	TGCAATTAATGCATTAAGCGATCTAGACAGATGCTGTAAAGACAGTGTGCTCCCAATATTC	1719
Db	707	TCACCTAATGATTAACAATTTTTTTTTTTTATGCTATTTTATATACACATATGTTAATTTA	6488
OY	172	AATCTAGCTCTTGCAATTTGGTGGATACATAATGTTATTAAGTGGTTTAAAGAAATCTTTAT	2319
Db	647	TCTGTAGCTTTCATTAACCTCTTCTTCAATTAACGTTTTCTTATTTTCAAAAGATGCTGTA	5888
OY	232	TCTGTAGGTAGAGTTAAAGATGCTTTATGATATATCCATGGGTGATTTCTTAAGTATATA	2919
Db	587	AATTTGAAGAAGCTAAATAATTAAGTTTGAATCTTAATTTTAAAAAATCATCTGTAAATATCAATCA	5288
OY	292	GGGGTTTTTCGACGAAGATATAGTTGGATATACATTTCCATGTCAGATATGTTTTCTGTA	3519
Db	527	AAATGTTCCATATGTGATCATTTTACAAAATGTTACTATTTGTAGATGTATTACTTTA	4688
OY	352	TTCATTAATPACCATTTTTCATAGCTATATATCAATACCTTGACAAATCATGAGATAGCGTTCTA	4119
Db	467	TTATTAATTAATTAATTAATTAATTAATCCATTGATGATTAATATTAATTAACCTGTACAA	4088
OY	412	CTATATATA 420	
Db	407	ACATTTTTTA 399	

RESULT

X99599 standard; DNA; 615 BP

AC	X99599;	
XX		
DT	05-OCT-1999	(first entry)
XX		

DE Nucleic acid sequence from U. urealyticum.
XX
KW ureaplasma urealyticum; nucleic acid detection; infection; pathogen;
KW human urogenital tract; pregnancy; neonatal disease; drug therapy;
KW suppurative arthritis; ss.
XX

OS Ureaplasma urealyticum
yy

PN WO9939007-A1

PD 05-AUG-1999

PF 29-JAN-1999; 99WO-US01972.

PR 30-JAN-1998; 98US-0073189

PA (UABR-) UAB RES FOUND

PI Cassell GH, Chen EY, Glass JI, Glass JS, Heltner CR,

	XX	:
		D
		C
		X
		O
		A
		F
		E
		E

DN WFL, 1233-403343/33
XX

PI detection of reactivated ureaplasma using novel genes, probes and primers

xx
PS Claim 1: Page 71: 110bn: End]ish

CC The preart invention provides methods for the detection and diagnosis
CC of Ureaplasma urealyticum infection. It provides novel genes (X99501-681)
CC that can be used as a source of primers and probes for the detection and
CC or quantification of U. urealyticum in a biological sample. The probes
CC that can be used in the method of the invention by forming target:probe
CC complex is complementary to a region selected from one of the 181
CC nucleic acid sequences (X99501-681). U. urealyticum is an opportunistic
CC pathogen of the human urogenital tract that is a significant cause of
CC adverse pregnancy outcome, neonatal disease, and suppurative arthritis.
CC As the infections are commonly asymptomatic, it is important to have
CC specific and sensitive methods for detecting their presence in a patient.
CC Also, as the pathogen has no current antibiotic directed specifically
CC against it, it would be advantageous to isolate and detect gene sequences
CC which are unique to it, and utilise these as a basis for diagnosis of
CC U. urealyticum infection as well as to develop new and improved drug
CC therapies. The present invention provides such novel polynucleotide
CC sequences (X99501-681).

Sequence 615 BP; 227 A; 64 C; 60 G; 264 T; 0 other;

	Score	DB	Length
Query Match	7.18	39.2	615
Best Local Similarity	45.38		
Matches 183; Conservative	0; Mismatches	218;	Indels 3; Gaps 1;

[illegible]

[illegible]

Db	1561	gtctaaatttttttctatgctgtttaacgttacttaacgtlatacttaatgctgtttaataatg	1620
Qy	255	TTTTAATCAATCATCGATCGGTATTTCTTAACCTTAATAGGGGTTTTGCAGCAAGATATGG	314
Db	1621	ttctaaagtcttctacgttacttaacgtcacttaacgttcttaagtcttgaagta	1680
Qy	315	TTGATATCATCTTTCCTAGTGTCAAGTATTTGTTTTCTTATCAATTAATAGCATATTTTCATAGC	374
Db	1681	taggaataaaaaatagaattttttaaatttttttctaatcttgtaaaaaataataat	1740
Qy	375	TATATCATATCTTGAC	390
Db	1741	gattcaacatgtygc	1756
RESULT	7		
ID	V71738		
XX	V71738	standard; DNA; 5407 BP.	
AC	V71738;		
XX	08-FEB-1999	(first entry)	
DT			
XX			
DE		Upstream sequence of the ACC synthase gene.	
XX			
KW		Apple beta galactosidase; ABG1; ACC; promoter; ripening; fruit; apple;	
XX		1-Aminocyclopropane-1-Carboxylate synthase; transgenic; genome mapping;	
KW		transcription factor; ss.	
XX			
XX		Malus sylvestris.	
OS			
XX			
PN		WO9845445-A1.	
XX			
PD		15-OCT-1998.	
XX			
PE		03-APR-1998; 98WO-GB01000.	
XX			
PR		31-MAY-1997; 97GB-0011233.	
PR		09-APR-1997; 97GB-0007193.	
XX			
PA		(UKAG-) UK MIN FISHERIES & FOOD.	
XX			
PI		Gillins JR, Hiles ER, James DJ;	
XX			
DR		WPI; 1998-583199/49.	
XX			
PT		New inducible plant promoters from apple - useful to construct	
XX		vectors for tissue-specific expression of transgenes	
PS		Disclosure; Fig 6; 52pp; English.	
XX			
CC		This represents the upstream region of the 1-Aminocyclopropane-1-	
CC		Carboxylate synthase (ACC) gene and incorporates a promoter sequence.	
CC		The invention provides a new recombinant polynucleotide that comprises a	
CC		promoter sequence which is an inducible promoter obtainable from apple or	
CC		a functional portion of that promoter. The promoter is most preferably	
CC		the Apple beta Galactosidase (ABG1) promoter or the ACC synthase promoter	
CC		and is activated in response to tissue specific agents (especially agents	
CC		specific to ripening fruits). A host plant cell transformed with a	
CC		replication vector comprising one of the polynucleotides and a	
CC		replication element which permits replication of the vector can be used	
CC		in a method to produce transgenic apple plants. The transgenic plants can	
CC		be cultivated to produce apples having a modified phenotype. The	
CC		invention may be used to induce protein expression in specific plant	
CC		tissue, for example in order to control fruit ripening, or to introduce	
CC		protective proteins. The invention may also be used to isolate other	
CC		ripening-specific promoters or transcription factors, or in genome	
CC		mapping. A probe comprising one of the promoter sequences can be used	
CC		to separate a transcription factor from fruit cells.	
XX			
SO		Sequence 5407 BP; 1828 A; 930 C; 854 G; 1795 T; 0 other;	
XX			

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 30, 2001, 21:48:22 ; Search time 60.87 Seconds
(without alignments)
1591.991 Million cell updates/sec

Title: US-09-382-242-23

Perfect score: 555
Sequence: 1 AAGCTTTTAACACAGCCTC.....AATCATACACTAAGCATAG 555

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 302621 segs, 87301344 residues

Total number of hits satisfying chosen parameters: 605242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_MA:*

1: /cgnl_7/ptodata/1/ina/5A.COMB.seq:*
2: /cgnl_7/ptodata/1/ina/5B.COMB.seq:*
3: /cgnl_7/ptodata/1/ina/6A.COMB.seq:*
4: /cgnl_7/ptodata/1/ina/6B.COMB.seq:*
5: /cgnl_7/ptodata/1/ina/PCUS.COMB.seq:*
6: /cgnl_7/ptodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	555	100.0	555	2	US-08-602-359A-23
2	38	6.8	152331	4	US-09-128-155-16
3	38	6.8	176373	4	US-09-128-155-17
4	36	6.5	4673	1	US-07-638-431-1
5	36	6.5	4673	5	PCT-US92-00018-1
6	35.6	6.4	1218	2	US-08-731-722-4
7	34.8	6.3	1958	1	US-08-137-175A-7
8	34.8	6.3	1958	4	US-08-479-017-7
9	34	6.1	5775	1	US-08-306-691B-15
10	34	6.1	5775	5	PCT-US93-06251-29
11	33.6	6.1	446	4	US-09-097-541-1
12	33.6	6.1	19124	2	US-08-487-826B-13
13	33.2	6.0	450	1	US-08-090-523-28
14	33.2	6.0	450	1	US-08-398-627-28
15	33.2	6.0	450	1	US-08-406-857-2
16	33.2	6.0	450	1	US-08-596-024-4
17	33.2	6.0	450	5	PCT-US94-07072-2
18	33	5.9	1853	1	US-07-849-438-2
19	32.8	5.9	2014	2	US-08-525-742-7
20	32.8	5.9	2022	2	US-08-505-486-96
21	32.8	5.9	2022	3	US-08-801-028-96
22	32.8	5.9	2022	3	US-09-340-154-96
23	32.8	5.9	2022	5	PCT-US95-09338-96
24	32.8	5.9	2022	5	PCT-US95-09339-96
25	32.8	5.9	2127	2	US-08-505-486-95
26	32.8	5.9	2127	2	US-08-801-028-95
27	32.8	5.9	2127	3	US-09-340-154-95

C 28	32.8	5.9	2127	5	PCT-US95-09338-95	Sequence 95, Appl
C 29	32.8	5.9	2127	5	PCT-US95-09339-95	Sequence 95, Appl
C 30	32.6	5.9	5910	1	US-08-195-814-1	Sequence 1, Appl
C 31	32.4	5.8	1186	2	US-08-731-722-5	Sequence 5, Appl
C 32	32.4	5.8	1617	2	US-08-467-963C-25	Sequence 25, Appl
C 33	32.4	5.8	1617	2	US-08-838-189D-25	Sequence 25, Appl
C 34	32.4	5.8	1617	3	US-08-852-344D-25	Sequence 25, Appl
C 35	32.4	5.8	1617	3	US-08-344-639E-25	Sequence 25, Appl
C 36	32.4	5.8	1844	2	US-08-467-963C-1	Sequence 25, Appl
C 37	32.4	5.8	1844	2	US-08-838-189D-1	Sequence 1, Appl
C 38	32.4	5.8	1844	3	US-08-852-344D-1	Sequence 1, Appl
C 39	32.4	5.8	1844	3	US-08-344-639E-1	Sequence 1, Appl
C 40	32.4	5.8	1844	4	US-08-467-969A-1	Sequence 1, Appl
C 41	32.4	5.8	1844	4	US-08-467-961A-1	Sequence 1, Appl
C 42	32.2	5.8	5181	1	US-08-257-073-10	Sequence 10, Appl
C 43	32	5.8	1697	1	US-08-472-028A-9	Sequence 9, Appl
C 44	32	5.8	1697	1	US-09-071-286-9	Sequence 9, Appl
C 45	32	5.8	5117	4	US-08-834-385-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-602-359A-23
Sequence 23, Application US/08602359A
Patent No. 5942430
GENERAL INFORMATION:
APPLICANT: ROBERTSON, Daniel E.
APPLICANT: MOREHY, Dennis
APPLICANT: REID, John
APPLICANT: MAFFIA, Anthony
APPLICANT: LINK, Steven
APPLICANT: SWANSON, Ronald V.
APPLICANT: WARREN, Patrick V.
APPLICANT: KOSMOTKA, Anna
TITLE OF INVENTION: ESTERASES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESS: 4225 EXECUTIVE SQUARE, STE 1400
CITY: LA JOLLA
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/602,359A
FILING DATE: February 16, 1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: HALL, LISA A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/010001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-678-5070
TELEFAX: 619-678-5099
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 555 NUCLEOTIDES
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: GENOMIC DNA
US-08-602-359A-23

Query Match 100.0%; Score 555; DB 2; Length 555;
Best Local Similarity 100.0%; Pred. No. 3.5e-137;
Matches 555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ATGCTTTAAACAGCAGCTTTGGATGATATGATATTTATTCACGTTTCTTTC 60
DB 1 ATGCTTTAAACAGCAGCTTTGGATGATATGATATTTATTCACGTTTCTTTC 60
QY 61 CCAATTAACATGATCCGATAGCATCTATGCGTCATGCTTATATATGGAATAT 120
DB 61 CCAATTAACATGATCCGATAGCATCTATGCGTCATGCTTATATATGGAATAT 120
QY 121 GCATTAAGCATCTAGACATGCTGTTAAAGAGTGTCTCCAAATATTCATCTAGT 180
DB 121 GCATTAAGCATCTAGACATGCTGTTAAAGAGTGTCTCCAAATATTCATCTAGT 180
QY 181 CTTCGCAATTTGGGATGATATGATATGATGTTAAGAAATCTTATCTGGAGT 240
DB 181 CTTCGCAATTTGGGATGATATGATATGATGTTAAGAAATCTTATCTGGAGT 240
QY 241 AGAGTTAAAGATCTTATCATATCCATGCGTATTTCTTAACCTAATAGGGGTTTC 300
DB 241 AGAGTTAAAGATCTTATCATATCCATGCGTATTTCTTAACCTAATAGGGGTTTC 300
QY 301 GACGAGATATATGTTGGATACATTTCTAGTCTGATGATGTTTCTATCAATATA 360
DB 301 GACGAGATATATGTTGGATACATTTCTAGTCTGATGATGTTTCTATCAATATA 360
QY 361 GCATTTTCTAGCTATATCAATATGATGATGATGATGATGATGATGATGATGAT 420
DB 361 GCATTTTCTAGCTATATCAATATGATGATGATGATGATGATGATGATGATGAT 420
QY 421 GGTCAATATTCGCAATGTTGATCTACACTTGTCTGACAGATTCGAGAGGTCGAT 480
DB 421 GGTCAATATTCGCAATGTTGATCTACACTTGTCTGACAGATTCGAGAGGTCGAT 480
QY 481 CCCGAGTATTTAGCGTATTTCTGTTTACATCTATATTAAGACGATATTTAAATCA 540
DB 481 CCCGAGTATTTAGCGGATTTCTGTTTACATCTATATTAAGACGATATTTAAATCA 540
QY 541 TACACTAAACGATAG 555
DB 541 TACACTAAACGATAG 555
```

RESULT 2

US-09-128-155-16/C
; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; EARLIER FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 152331
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(152331)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

Query Match 6.8%; Score 38; DB 4; Length 152331;
Best Local Similarity 53.3%; Pred. No. 0.77;
Matches 80; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

```
QY 234 GTGAGTAGAGCTTAAGAGATCTTAAATCAATCCATGCGTATTTCTTAACCTAATAGG 293
DB 137980 GTGAGTAGAGCTTCACTCTTCAATATGATGATGATGATGATGATGATGATGATG 137921
QY 294 GCTTTTGACGACGATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 353
DB 137920 TGTGTCGACAAATATTTGTAAGATACATTTCTCACTTAATGATGATGATGATGATG 137861
QY 354 AATAATAGCATATTTCAATAGCTATATCAT 383
DB 137860 CTCAATATGATATCTCAAAACATGCTTCAT 137831
```

RESULT 3

US-09-128-155-17/C
; Sequence 17, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; EARLIER FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 176373
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(176373)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-17

Query Match 6.8%; Score 38; DB 4; Length 176373;

Best Local Similarity 53.3%; Pred. No. 0.8; Matches 80; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

```
QY 234 GTGAGTAGAGCTTAAGAGATCTTAAATCAATCCATGCGTATTTCTTAACCTAATAGG 293
DB 150561 GTGAGTAGAGCTTCACTCTTCAATATGATGATGATGATGATGATGATGATGATG 150502
QY 294 GCTTTTGACGACGATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 353
DB 150501 TGTGTCGACAAATATTTGTAAGATACATTTCTCACTTAATGATGATGATGATGATG 150442
QY 354 AATAATAGCATATTTCAATAGCTATATCAT 383
DB 150441 CTCAATATGATATCTCAAAACATGCTTCAT 150412
```

RESULT 4

US-07-638-431-1
; Sequence 1, Application US/07638431
; Patent No. 5198535
; GENERAL INFORMATION:
; APPLICANT: Hoffman, Stephen L.
; APPLICANT: Charoenvit, Yupin
; APPLICANT: Hegstrom, Richard
; APPLICANT: Khushf, William O.
; APPLICANT: Rogers IV, William O.

RESULT 5
 PCT-US92-00018-1
 : Sequence 1, Application PC/TUS9200018
 : GENERAL INFORMATION:
 : APPLICANT: Hoffman, Stephen L.
 : APPLICANT: Charoenvit, Yupin
 : APPLICANT: Hedstrom, Richard
 : APPLICANT: Khumsmith, Srisin
 : APPLICANT: Rogers IV, William O.
 : TITLE OF INVENTION: Protective malaria sporozoite surface protein

```

1  RESULT 6
2  US-08-731-722-4
3  : Sequence 4, Application US/08731722
4  : Patent No. 5961971
5  :
6  : GENERAL INFORMATION:
7  :
8  : APPLICANT: Martii, Frank N.
9  :
10 : TITLE OF INVENTION: Biocontrol of Fungal Soilborne Pathogens
11 :
12 : TITLE OF INVENTION: by Pythium oligandrum
13 :
14 : NUMBER OF SEQUENCES: 8
15 :
16 : CORRESPONDENCE ADDRESS:
17 :
18 : ADDRESSEE: Saliwanchik & Saliwanchik
19 :

```

STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: US
ZIP: 32606-6669
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/731,722
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Whitlock, Ted W.
REGISTRATION NUMBER: 36,965
REFERENCE/DOCKET NUMBER: UF-161
TELECOMMUNICATION INFORMATION:
TELEPHONE: 352-375-8100
TELEFAX: 352-372-5800
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: 17-1
US-08-731-722-4

Query Match 6.4%; Score 35.6; DB 2; Length 1218;
Best Local Similarity 48.1%; Pred. No. 0.8;
Matches 101; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 150 AAGCAGTGTGCTCCAAATTCATCTGAGTTCGCAATGGGGATCAATATGTTAT 209
DB 512 AACTTGTCTAAATCCACCAAAATTTAACTATATTAATGAAAGACTTTTACTTAT 571
QY 210 AGTTGGTTAAGAAATTTATCGTGAGTAGAGTTAAGATCTTAATCATATCCAT 269
DB 572 ATTCTTAATAATATCTGAATTTAAACAGAGACCTTAAGACCTTAATAATAGCTAT 631
QY 270 GGGTATTTCTTAACCTAATAGGGCTTTTCGACGAGATATGTTGATACATTTCT 329
DB 632 AAGCAATTTTTCATAGAAAGSTTTATTTTAAATGAAATTTGTAATCTTATTT 691
QY 330 AGTCTCAGTATGTTTTCCTTATCATATAT 359
DB 692 TACTTTAATAAGAAATTAATTAATTAATAT 721

RESULT 7
US-08-137-175A-7
Sequence 7, Application US/08137175A
Patent No. 5777095
GENERAL INFORMATION:
APPLICANT: BARBOUR, Alan G.
APPLICANT: BERGSTROM, Sven
APPLICANT: HANSSON, Lennart
TITLE OF INVENTION: IMPROVEMENT IN BORRELIA BURGDOFFERI AND
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/137,175A
FILING DATE: 26-Oct-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/08972
FILING DATE: 22-Oct-1992
ATTORNEY/AGENT INFORMATION:
NAME: COOPER, Iver P.
REGISTRATION NUMBER: 28,005
REFERENCE/DOCKET NUMBER: BARBOUR-1B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Borrelia burgdorferi
STRAIN: 1p90
INDIVIDUAL ISOLATE: Isolate from I. persulcatus from
INDIVIDUAL ISOLATE: Soviet Union
FEATURE:
NAME/KEY: CDS
LOCATION: 125..949
OTHER INFORMATION: /product= "OspsA"
NAME/KEY: CDS
LOCATION: 959..1843
OTHER INFORMATION: /product= "OspsB"
US-08-137-175A-7

Query Match 6.3%; Score 34.8; DB 1; Length 1958;
Best Local Similarity 48.5%; Pred. No. 1.5;
Matches 96; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 196 ATACTAATTTGTTATAGTTCGTTAAGAAATCTTATTCGCTGAGTAGATTAAGATCT 255
DB 5 ATATTGAAATTAATTAATCATTTTATTTTATTTTATTTGCTATTTGTTATTTGCT 64
QY 256 TTAATCATATCCATGCGTATTTCTTAACCTAATAGGGGTTTCGACGAGATATAGT 315
DB 65 TATACATAATTAATATTTGTTATTAAGTTATATTAATTAATTAATTAATTAATTA 124
QY 316 TGGATACATTTCTAGTCTCGATTTGTTTCTTAATCATTAATAGCATATTTATAGCT 375
DB 125 ATGAATAAATTAATTTAGGGAATAGCTAATTAATTAATTAATTAATTAATTAAT 184
QY 376 ATATCAATCTTGACAAA 393
DB 185 GTTAGACCTTGATGAA 202

RESULT 8
US-08-479-017-7
Sequence 7, Application US/08479017
Patent No. 6143872
GENERAL INFORMATION:
APPLICANT: BARBOUR, Alan G.
APPLICANT: BERGSTROM, Sven
APPLICANT: HANSSON, Lennart
TITLE OF INVENTION: IMPROVEMENT IN BORRELIA BURGDOFFERI AND
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:


```
RESULT 10
PCT-US93-06251-29
; Sequence 29, Application PC/TUS9306251
; GENERAL INFORMATION:
; APPLICANT: Mickstrom, Eric and Rife, Jason P.
; TITLE OF INVENTION: Trivalent Synthesis of Oligonucleotides Containing
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; City: Garden City
; STATE: NY
; COUNTRY: USA
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/06251
; FILING DATE: 19930630
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Digilio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 8586
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-742-4343
; TELEFAX: 516-742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5775 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; PCT-US93-06251-29

Query Match
Best Local Similarity 6.1%; Score 34; DB 5; Length 5775;
Matches 130; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

QY 4 TCTTTAAACAGCAGCTCTGGATGATGATATATTTATCTCAGCTTTCTTCCCA 63
DB 4538 TTGTCTCAAAAAGTAATATTTCTTGATATTTGTAGTATGTTTAAACCCAGCA 4597
QY 64 TTAACAATGATCCGATTAGCTATCTCTATGTCGATGGTTTAAATATGGAATATGCA 123
DB 4598 GTTAACCTTGAAAGCTGAATTTATTTAGTAACCTTCTGTATTAAGTATGATGAA 4657
QY 124 TTAAGCATCTAGACATGCTGTTTAAAGCAGTGTCTCCATATTCATCTAGCTT 183
DB 4658 TTCTGCAATGAGAAAGCAATAGCTGTCATTAATGCTTTCTTCCAAAGAAAGTACT 4717
QY 184 GCAATTCGTGGGATCTAATGTTTATGTTAGTAAAGAAATCTTTATGCTGAGTGA 243
DB 4718 CACATGAGTCTTGAAGAATATGATATAGTAAAGTATGATGCTGTTTAAATAGT 4777
QY 244 GTTAAAGCATCTTAAATCATATCCATGCGGTGATTTCTTAACTTAATAGC 293
DB 4778 TTGAAGTGCTGTTGGATATGATAGTAAATTTAGATCAATTTAGGGG 4827

RESULT 11
US-09-097-541-1
; Sequence 1, Application US/09097541
; Patent No. 6100092
; GENERAL INFORMATION:
; APPLICANT: Borisyuk, Mykola
```

```
APPLICANT: Borisyuk, Lyudmyla
; APPLICANT: Raskin, Ilya
; TITLE OF INVENTION: Materials and methods for amplifying
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive/6300 Sears Tower
; City: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/097,541
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Merkel, William K.
; REGISTRATION NUMBER: 40,725
; REFERENCE/DOCKET NUMBER: 29155/34699
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 446 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "aps rdna sequence"
; US-09-097-541-1

Query Match
Best Local Similarity 6.1%; Score 33.6; DB 4; Length 446;
Matches 114; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY 196 ATACTAATGTTAATAGTGGTTTAAAGAAATCTTTATTCGTGAGTAGATTAAGATCT 255
DB 126 ATTTATGATGATGTTGATTTTAAATAGTTTATGCTGTTTAAATATTTT 185
QY 256 TTAATCATATCCATGGGATTTCTTAATATAGGGTTTCGACGAATATAGT 315
DB 186 ATTTATTTATGTTTAAATATTTAATTTAGATGATTTGGAATTTCTCCATGTTTCT 245
QY 316 TGATATATTCCTAGTCTCAGTATGCTTTTCTTATCAATATAGCATATTTTCATAGCT 375
DB 246 ATATTATTAATATTTCTTATTTTGTATTTATATATATTTTGTGTTTATAT 305
QY 376 ATATCAATCTTGACAAATCATGATAGCTGTTCTACTAATATAGTCATATTCGATG 435
DB 306 AAATATTTATTAATAAAATATTTATTTTGTAAATAATATATCATTTACAAATGTTTAAAG 365
QY 436 TGTATCT 443
DB 366 TCATTTGT 373

RESULT 12
US-08-487-826B-13/C
; Sequence 13, Application US/08487826B
; Patent No. 593827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chilens, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
```


APPLICANT: Stark, David M.
TITLE OF INVENTION: Enhanced Starch Biosynthesis
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: Grace L. Bonner, Monsanto Co. BBAF
STREET: 700 Chesterfield Parkway No. 5608149th
CITY: St. Louis
STATE: Missouri
COUNTRY: USA
ZIP: 63198
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/398,627
FILING DATE: 03-MAR-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/090,523
FILING DATE: 12-JUL-1993
APPLICATION NUMBER: US 07/709663
FILING DATE: 07-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/539763
FILING DATE: 18-JUN-1990
ATTORNEY/AGENT INFORMATION:
NAME: Bonner, Grace L.
REGISTRATION NUMBER: 32,963
REFERENCE/DOCKET NUMBER: 38-21(10559)A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (314) 537-7286
TELEFAX: (314) 537-6047
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 450 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-398-627-28

Query Match 6.08; Score 33.2; DB 1; Length 450;
Best Local Similarity 50.08; Pred. No. 2.6;
Matches 83; Conservative 0; Mismatches 83; Indels 0; Gaps 0;
QY 277 TTCTTAACTTAATAGGGCTTTTCGACGAATATATGTTGATACATTTCTAGTCTCA 336
||||| ||||| - - - - - ||||| ||| - - -
DB 48 TTCTTTTAAATGATTTCTTCTACTCTAATTTGAACATATTTGAGACCACTTAATGTA 107
QY 337 GTATTGTTTTCTTATCAATATAGCATATTTTCATAGCTATCAATCTTGACAAATCA 396
- - - - - ||||| ||||| - - - - - ||||| ||| - - -
DB 108 TTTTCTTTTCTTATCAAAATGATGCTCTATATTAATATCAATGTTATATACA 167
QY 397 TCGATAGCTGTTCTACTAATAATAGTCATATTTGCAATGTGATC 442
- - - - - ||||| ||||| - - - - - ||||| ||| - - -
DB 168 TAAATTTAATATTTTATATAAAAAAATATCGAGCTAATCATATC 213

RESULT 15
US-08-406-857-2
Sequence 2, Application us/08406857
Patent No. 5608150
GENERAL INFORMATION:
APPLICANT: Conner, Timothy W.
TITLE OF INVENTION: Fruit Specific Promoters
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Grace L. Bonner, Monsanto Company, BBAF
STREET: 700 Chesterfield Parkway No. 5608150th
CITY: St. Louis

STATE: Missouri
COUNTRY: USA
ZIP: 63198
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/406,857
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/07072
FILING DATE: 27-JUN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/090,523
FILING DATE: 12-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Bonner, Grace L.
REGISTRATION NUMBER: 32,963
REFERENCE/DOCKET NUMBER: 38-21(10655)A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (314)537-7286
TELEFAX: (314)537-6047
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 450 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-406-857-2

Query Match 6.08; Score 33.2; DB 1; Length 450;
Best Local Similarity 50.08; Pred. No. 2.6;
Matches 83; Conservative 0; Mismatches 83; Indels 0; Gaps 0;
QY 277 TTCTTAACTTAATAGGGCTTTTCGACGAATATATGTTGATACATTTCTAGTCTCA 336
||||| ||||| - - - - - ||||| ||| - - -
DB 48 TTCTTTTAAATGATTTCTTCTACTCTAATTTGAACATTTGAGACCACTTAATGTA 107
QY 337 GTATTGTTTTCTTATCAATATAGCATATTTTCATAGCTATCAATCTTGACAAATCA 396
- - - - - ||||| ||||| - - - - - ||||| ||| - - -
DB 108 TTTTCTTTTCTTATCAAAATGATGCTCTATATTAATATCAATGTTATATACA 167
QY 397 TCGATAGCTGTTCTACTAATAATAGTCATATTTGCAATGTGATC 442
- - - - - ||||| ||||| - - - - - ||||| ||| - - -
DB 168 TAAATTTAATATTTTATATAAAAAAATATCGAGCTAATCATATC 213

Search completed: May 30, 2001, 22:40:51
Job time: 3149 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 30, 2001, 21:07:06 ; Search time 1108.56 Seconds
(without alignments)
4373.723 Million cell updates/sec

Title: US-09-382-242-23

Perfect score: 555

Sequence: 1 ATGCTTTTAAACAAGCACTC.....AATCATACACTTAAACGATAG 555

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 9623517 seqs, 4368049070 residues

Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: gb_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estp11:*
73: em_estp12:*
74: em_estp13:*
75: em_estp14:*
76: em_estp15:*
77: em_estp16:*
78: em_estp17:*
79: em_estp18:*
80: em_estp19:*
81: em_estp110:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
 118: gb_est49:*
 119: gb_est50:*
 120: gb_est51:*
 121: gb_est52:*
 122: gb_est53:*
 123: gb_est54:*
 124: gb_est55:*
 125: gb_est56:*
 126: gb_est57:*
 127: gb_est58:*
 128: gb_est59:*
 129: gb_est60:*
 130: gb_est61:*
 131: gb_est62:*
 132: gb_est63:*
 133: gb_est64:*
 134: gb_est65:*
 135: gb_est66:*
 136: gb_est67:*
 137: gb_est68:*
 138: gb_est69:*
 139: gb_est70:*
 140: gb_est71:*
 141: gb_est72:*
 142: gb_est73:*
 143: gb_est74:*
 144: gb_est75:*
 145: gb_est76:*
 146: gb_est77:*
 147: gb_est78:*
 148: gb_est79:*
 149: gb_est80:*
 150: gb_est81:*
 151: gb_est82:*
 152: gb_est83:*
 153: gb_est84:*
 154: gb_est85:*
 155: gb_est86:*
 156: gb_est87:*
 157: gb_est88:*
 158: gb_est89:*
 159: gb_est90:*
 160: gb_est91:*
 161: gb_est92:*
 162: gb_est93:*
 163: gb_est94:*
 164: gb_est95:*
 165: gb_est96:*
 166: gb_est97:*
 167: gb_est98:*
 168: gb_est99:*
 169: gb_est100:*
 170: gb_est101:*
 171: gb_est102:*
 172: gb_est103:*
 173: gb_est104:*
 174: gb_est105:*
 175: gb_est106:*
 176: gb_est107:*
 177: gb_est108:*
 178: gb_est109:*
 179: gb_est110:*
 180: gb_est111:*
 181: gb_est112:*
 182: gb_est113:*
 183: gb_est114:*
 184: gb_est115:*
 185: gb_est116:*
 186: gb_est117:*
 187: gb_est118:*
 188: gb_est119:*
 189: gb_est120:*

190: em_gss_pln1:*
 191: em_gss_pln2:*
 192: em_gss_pro:*
 193: em_gss_rdd1:*
 194: em_gss_rdd2:*
 195: em_gss_rdd3:*
 196: em_gss_rdd4:*
 197: em_gss_rdd5:*
 198: em_gss_vrt1:*
 199: em_gss_vrt2:*
 200: em_gss_vrt3:*
 201: em_gss_vrt4:*
 202: em_gss_vrt5:*
 203: em_gss_vrt6:*
 204: em_gss_vrt7:*
 205: em_gss_vrt8:*
 206: em_gss_vrt9:*
 207: em_gss_vrt10:*
 208: em_gss_vrt11:*
 209: em_gss_vrt12:*
 210: em_gss_vrt13:*
 211: em_gss_vrt14:*
 212: em_gss_vrt15:*
 213: em_gss_vrt16:*
 214: em_gss_vrt17:*
 215: em_gss_vrt18:*
 216: em_gss_vrt19:*
 217: em_gss_vrt20:*
 218: em_gss_vrt21:*
 219: em_gss_vrt22:*
 220: em_gss_vrt23:*
 221: em_gss_vrt24:*
 222: em_gss_vrt25:*
 223: em_gss_vrt26:*
 224: em_gss_vrt27:*
 225: em_gss_vrt28:*
 226: em_gss_vrt29:*
 227: em_gss_vrt30:*
 228: em_gss_vrt31:*
 229: em_gss_vrt32:*
 230: em_gss_vrt33:*
 231: em_gss_vrt34:*
 232: em_gss_vrt35:*
 233: em_gss_vrt36:*
 234: em_gss_vrt37:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	49.6	8.9	1101	CNS00EVL	AL069706 Drosophila
2	47.4	8.5	1101	CNS00YWL	AL096927 Drosophila
3	46.6	8.4	1101	CNS000D1	AL065414 Drosophila
4	46.6	8.4	1101	CNS0039G	AL063921 Drosophila
5	46.2	8.3	878	CNS0187R	AL108993 Drosophila
6	45.4	8.2	500	AU086529	AU086529 Drosophila
7	45.4	8.2	792	BF616219	BF616219 HVSMEC000
8	45.4	8.2	1101	CNS0106X	AL098595 Drosophila
9	45.2	8.1	1101	CNS0161L	AL106896 Drosophila
10	45	8.1	641	AQ946120	AQ946120 Sheared D
11	44.8	8.1	982	CNS013M1	AL103027 Drosophila
12	44.8	8.1	988	CNS0072R	AL066743 Drosophila
13	44.6	8.0	491	FR0042156	AL129648 Fugu rubr
14	44.6	8.0	502	FR0042075	AL129647 Fugu rubr
15	44.4	8.0	500	AU086530	AU086530 Drosophila
16	44.4	8.0	856	AZ687828	AZ687828 ENTLE61TR
17	44.4	8.0	911	AZ677405	AZ677405 ENTJN1TR
18	44.4	8.0	1146	CNS021G2	AL176843 Tetradon

BASE COUNT	419 a	91 c	60 g	299 t	232 others
ORIGIN	/note="end : T7"				
Query Match	8.9%; Score 49.6; DB 229; Length 1101;				
Best Local Similarity	32.7%; Pred. No. 0.031;				
Matches 148; Conservative	70; Mismatches 234; Indels 0; Gaps 0;				
QY	103	TTTATATATGAAATTAATGATTAAGCATGATGAGACATGCTGTTAAAGCAGTGCTGCT	162		
Db	963	TWMTATATMTATTAATTAATAATATATATATATATATATATATATATATATATATATATAT	904		
QY	163	CAAATATTCCTTAACTGATCTTGCATTTGGTGAGATACATAATTTAGTTAGTTTAAACA	222		
Db	903	ATAATWTTWMAAAAMWMAATMTATATTTTWTATATATATATATATATATATATATATAT	844		
QY	223	AATCTTATATCGTGAGATGAGATTAAGGATCTTAATCATATATCATATGATGCTATTTCTT	282		
Db	843	TWMTATATMTATMTATMTATMTATMTATMTATMTATMTATMTATMTATMTATMTATMTATA	784		
QY	283	AACCTAATAGGGGTTTTGACGAAGTATATGTTGGATACATTTCTAGTCACAGATATG	342		
Db	783	MMMTTMTATMTTATAAAMATATATATMTATATATATATATATATATATATATATATAT	724		
QY	343	TTTTTTCTTATCATTAATGCAATATTTTCATATAGCTATATATACAAATACATGACAAATCATGATA	402		

[illegible]

```

FEATURES
  source      location/qualifiers
1. 1101      /organism="Drosophila melanogaster"

```

/plasmid="pBelOBAC11"
/db_xref="taxon:7227"
/clone_11b="DrosBAC"
/clone="BACN01G13"
/note="end : SP6"
BASE COUNT 424 a 119 c 129 g 248 t 181 others
ORIGIN

Query Match 8.5%; Score 47.4; DB 229; Length 1101;
Best Local Similarity 35.1%; Pred. No. 0.11;
Matches 98; Conservative 46; Mismatches 135; Indels 0; Gaps 0;

OY 106 AATATATGGAATATGATTAAGCATCTAGACATGCTGTTAAAGCAGTGTGCTCA 165
DB 1100 HAAATATATATATATATATATATATATATATATATATATATATATATAT 1041
OY 166 AATATCAATCTAGCTCTTCCATATGCTGATACATATATGATATAGTGTGA 225
DB 1040 AATATATATATATATATATATATATATATATATATATATATATATATAT 981
OY 226 CTATATCTGAGTATAGATATAGATCTTAAATATATATATATATATATATAT 285
DB 980 AATATATATATATATATATATATATATATATATATATATATATATATAT 921
OY 286 TTAATAGGGGTTTTCAGACAGATATATATGTTGATATCTTCTACTGATTT 345
DB 920 TTAATATATATATATATATATATATATATATATATATATATATATATAT 861
OY 346 TTTCTATCATATATATATATATATATATATATATATATATATATATATAT 384
DB 860 TTTTATATATATATATATATATATATATATATATATATATATATATAT 822

RESULT 3

CNS000D1

LOCUS

DEFINITION

CNS000D1 1101 bp DNA GSS 03-JUN-1999
Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR01J16 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
AL065414
GI:4938827

ACCESSION

KEYWORDS

SOURCE

ORGANISM

Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazuhiro Osoegawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

Source

1. 1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"

/clone_11b="RPCI-98"
/clone="BACR01J16"
/note="end : TET3"
BASE COUNT 280 a 104 c 123 g 211 t 383 others
ORIGIN

Query Match 8.4%; Score 46.6; DB 229; Length 1101;
Best Local Similarity 14.0%; Pred. No. 0.17;
Matches 63; Conservative 203; Mismatches 185; Indels 0; Gaps 0;

OY 21 TTGATGATATGATATATATATATATATATATATATATATATATATATATAT 80
DB 647 KKKKRAATDPTATATATATATATATATATATATATATATATATATATATAT 706
OY 81 AACTATCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 140
DB 707 KKKKDDKPTKKKKKTATATATATATATATATATATATATATATATATATAT 766
OY 141 TCGTGTAAAGCAGTGTGCTCAATATATATATATATATATATATATATATAT 200
DB 767 GDRRTATGARGDKTKKKKTGDRKGGDGGDGGDGGDGGDGGDGGDGGDGGDGG 826
OY 201 AATGTGTATATGCTGCTTACGAATCTTATTCGTGAGTACAGTTAAAGATCTT 260
DB 827 TGCANDKADKKRKRKRKRKRKRKRKRKRKRKRKRKRKRKRKRKRKRKRKRKR 886
OY 261 CATATCAGGCTGATATATATATATATATATATATATATATATATATATATAT 320
DB 887 GADKKKKDDDDTADNDTKDTATATATATATATATATATATATATATATATAT 946
OY 321 ACATTTCTAGTCTGATATATATATATATATATATATATATATATATATATAT 380
DB 947 RKDGKRTAGARGADKKKKKDRDPAKATDGDKDTWDPTDPAKRRRRRRRRGR 1006
OY 381 AATGCTTGAATATGATATATATATATATATATATATATATATATATATATAT 440
DB 1007 GDRKGGKRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1066
OY 441 TCTACACTTGTCTGAGATTCGAGAGT 471
DB 1067 KAKAKDKAKKADDDKDKKAKAKAKDK 1097

RESULT 4

CNS0039G/c

LOCUS

DEFINITION

CNS0039G 1101 bp DNA GSS 03-JUN-1999
Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
AL063921
GI:4941778

ACCESSION

KEYWORDS

SOURCE

ORGANISM

Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazuhiro Osoegawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial

COMMENT Vector: pBluescript II KS

V_Type: phagemid
PRIMER: KS

DESCR: One pass dye-terminator sequencing of cosmid cloned genomic
sequence.

FEATURES Location/Qualifiers

source

1.502
/organism="Fugu rubripes"
/db_xref="taxon:31033"
/clone_lib="cosmid 151J19"
/clone="151J19C11"
BASE COUNT 210 a 76 c 76 g 123 t 17 others
ORIGIN

Query Match 8.0%; Score 44.6; DB 232; Length 502;
Best Local Similarity 46.6%; Pred. No. 0.48; Indels 0; Gaps 0;

Matches 125; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

QY 147 TAAAGAGTGTGCTCCAAATTCATCTAGCTCTGCAATGGGAGTACTAATGT 206

DB 475 TAGNAGNAGTATATATATATATAGTAGTATATATATATATATAGTAG 416

QY 207 TATAGTGTGTTAAGAAATCTTATTCGCGAGTAGAGTTAAGCATCTTATCATATC 266

DB 415 TAGNATTTATAGTATATATATATATATATATATATATATATATATATATAT 356

QY 267 CATGGGTATTTCTTAACTAATAGGGGTTTCGACGAGTATATGGTGTGATCATTT 326

DB 355 CGTTGNTATTCGATTTATTTATTTATTTGTCATTTAGATTAATGATTTGTTATTT 296

QY 327 CCTAGTCTAGTATGTTTCTTATCAATATATATATATATATATATATATATAT 386

DB 295 ACTATACGAGATATATCTGAGTATTTAGTGCATTAATATATATATATATAT 236

QY 387 TGACAAATCATGACATCTGCTCTACTA 414

DB 235 ATTTGATCATATTTATTTATTTGTTATTTA 208

RESULT 15
AU086530/c 500 bp mRNA EST 27-JAN-2001

LOCUS AU086530 Sugano Malaria cDNA library Plasmodium falciparum cDNA

DEFINITION Clone Xpfn2718, mRNA sequence.

ACCESSION AU086530

VERSION AU086530.1 GI:12388671

KEYWORDS EST.

SOURCE malaria parasite P. falciparum.

ORGANISM Plasmodium falciparum

REFERENCE Eukaryote: Alveolata: Apicomplexa: Haemosporida: Plasmodium.

AUTHORS 1 (bases 1 to 500)

TITLE Watanabe, J., Sasaki, M., Suzuki, Y., and Sugano, S.

JOURNAL Full-malaria: a database for a full-length enriched cDNA library

MEDLINE from human malaria parasite, Plasmodium falciparum

COMMENT Nucleic Acids Res. 29 (1), 70-71 (2001)

Contact: Junichi Watanabe
Institute of Medical Science
The University of Tokyo, Department of Parasitology
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan

Tel: 81-3-5449-5378

Fax: 81-3-5449-5410

Email: jwatanabe@med.s.u-tokyo.ac.jp

Suzuki, Y., Yoshitomo-Nakagawa, K., Maruyama, K., Suyama, A. and Sugano

S. Construction and characterization of a full length-enriched and

a 5'-end-enriched cDNA library Gene 200 (1-2), 149-156 (1997).

FEATURES

source

1.500
/organism="Plasmodium falciparum"
/strain="3D7"
/db_xref="taxon:5833"

/clone="Xpfn2718"
/clone_lib="Sugano Malaria cDNA library"
/dev_stage="erythrocytic stage"
BASE COUNT 244 a 58 c 47 g 149 t 2 others
ORIGIN

Query Match 8.0%; Score 44.4; DB 107; Length 500;

Best Local Similarity 48.6%; Pred. No. 0.54; Indels 0; Gaps 0;

Matches 120; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 295 GTTTTCGACGAGTATATGCTTGATACATTTCTAGTCTAGTATGTTTCTTATCA 354

DB 465 GTGTTTTTGAAGAAATATGCAATTTGCAATTTAATAATTTATTTATTTATTTA 406

QY 355 ATATAGCATATTTTCATGCTATATCATATCTGACAAATTCATGATGCTTTCTACTA 414

DB 405 TTATCATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 346

QY 415 ATATAGCATATTTTCATGCTATATCATATCTGACAAATTTCTAGTCTAGTATGCT 474

DB 345 TTATATCATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 286

QY 475 GCTATTTCCGAGTTATTTAGCGTATTTCTGTTTATTTACATTTCTATATATAGACATATTTT 534

DB 285 GATCTTCGAGTATTTTTCGAAATATTTGTTTACCATTTTATTTATTTATTTATTTT 226

QY 535 AATCAT 541

DB 225 TATCAT 219

Search completed: May 30, 2001, 22:38:17
Job time: 5471 sec

THIS PAGE BLANK (USPTO)